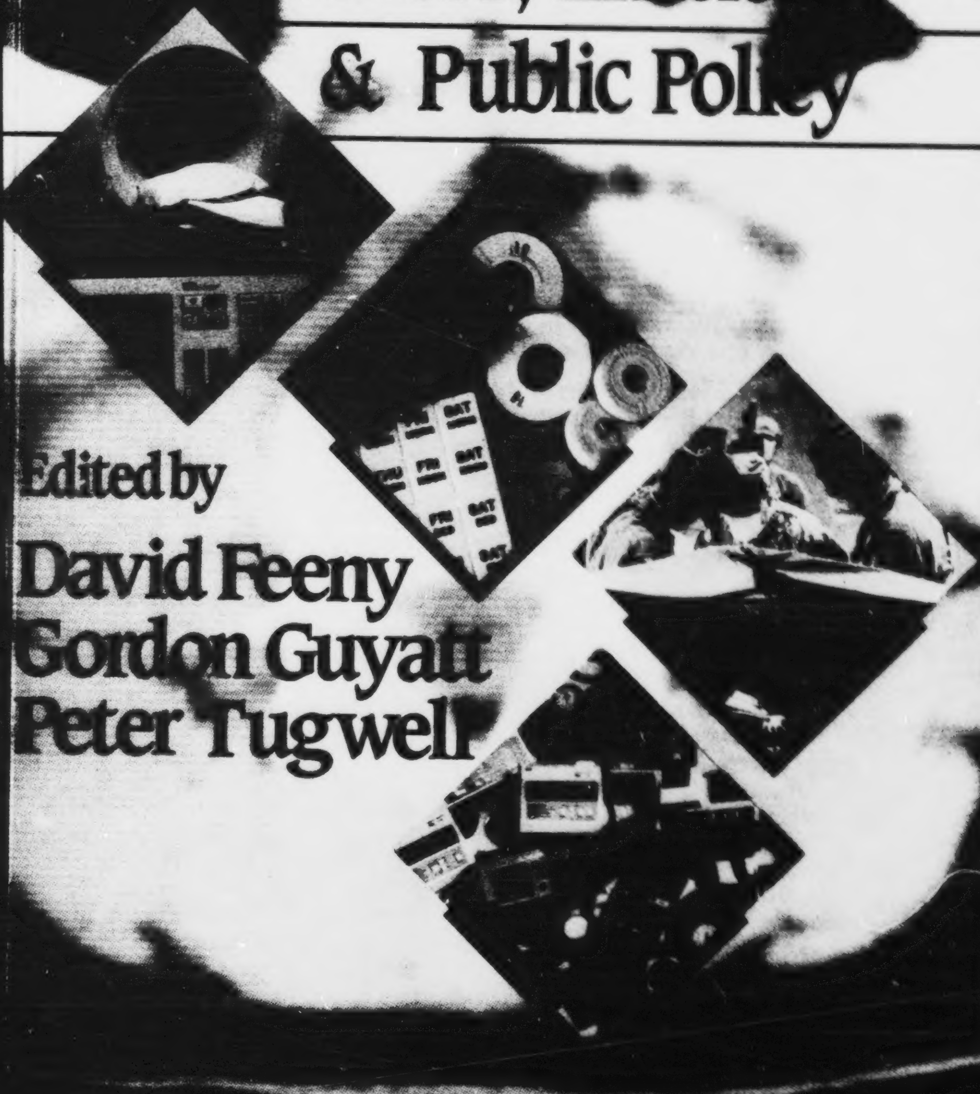
Effectiveness, Efficiency & Public Policy



Edited by

David Feeny
Gordon Guyatt
Peter Tugwell

86-06412/

The Institute for Research on Public Policy
L'Institut de recherches politiques

A national, independent, research organization
Un organisme de recherche national et indépendant

Founded in 1972, The Institute for Research on Public Policy is a national organization whose independence and autonomy are ensured by the revenues of an endowment fund which is supported by the federal and provincial governments and by the private sector. In addition, the Institute receives grants and contracts from governments, corporations and foundations to carry out specific research projects.

The *raison d'être* of the Institute is threefold:

- To act as a catalyst within the national community by helping to facilitate informed public debate on issues of major public interest
- To stimulate participation by all segments of the national community in the process that leads to public policy making
- To find practical solutions to important public policy problems, thus aiding in the development of sound public policies.

The Institute is governed by a Board of Directors, which is the decision-making body, and a Council of Trustees, which advises the Board on matters related to the research direction of the Institute. Day-to-day administration of the Institute's policies, programs and staff is the responsibility of the president.

The Institute operates in a decentralized way, employing researchers located across Canada. This ensures that research undertaken will include contributions from all regions of the country.

Wherever possible, the Institute will try to promote public understanding of, and discussion on, issues of national importance, with clarity and impartiality. Conclusions or recommendations in the Institute's publications are solely those of the author and should not be attributed to the Board of Directors, Council of Trustees, or contributors to the Institute.

The president bears final responsibility for the decision to publish a manuscript under the Institute's imprint. In reaching this decision, he is advised on the accuracy and objectivity of a manuscript by both Institute staff and outside reviewers. Publication of a manuscript signifies that it is deemed to be a competent treatment of a subject worthy of public consideration.

Publications of the Institute are published in the language of the author, along with an executive summary in both of Canada's official languages.

Health Care Technology:

Effectiveness, Efficiency, and Public Policy

Health Care Technology:

Effectiveness, Efficiency, and Public Policy

Edited by
David Feeny
Gordon Guyatt
Peter Tugwell

This study was jointly sponsored by:

The Canadian Medical Association
and
The Institute for Research on Public Policy

Copyright © The Institute for Research on Public Policy 1986
All rights reserved

Printed in Canada

Legal Deposit Second Quarter
Bibliothèque nationale du Québec

Canadian Cataloguing in Publication Data

Feeny, David, 1948-
Health Care technologies

Foreword and summary in English and French
ISBN 0-88645-038-1

1. Medical technology – Evaluation. 2. Medical technology – Government policy – Canada. I. Guyatt, Gordon, 1953- II. Tugwell, Peter III. Institute for Research on Public Policy IV. Title.

R856.F44 1986 610'.28 C86-090213-7

The Institute for Research on Public Policy/
L'Institut de recherches politiques
2149 Mackay Street,
Montreal, Quebec H3G 2J2

Contents

Foreword vii

Avant-Propos ix

Preface xiii

Acknowledgements xv

Contributors xvii

Summary xix

Abrégé xxi

1: Introduction: Health Care Technology 1
David Feeny

2: New Health Care Technologies: Their Effect on Health and the Cost of Health Care 5
David Feeny

3: The Diffusion of New Health Care Technologies 25
David Feeny

4: A Framework for the Evaluation of Technology: The Technology Assessment Iterative Loop 41
Peter Tugwell, Kathryn Bennett, David Feeny, Gordon Guyatt, R. Brian Haynes

5: **Guidelines for Health Technology Assessment: Therapeutic Technologies** 57
Gordon Guyatt, Michael Drummond, David Feeny, R. Brian Haynes, Peter Tugwell

6: **Guidelines for Health Technology Assessment: Diagnostic Technologies** 79
Gordon Guyatt, Michael Drummond, David Feeny, R. Brian Haynes, Peter Tugwell

7: **Guidelines for Health Technology Assessment: Economic Evaluation** 107
Michael Drummond

8: **Planning for the Provision and Utilization of New Health Care Technologies** 129
Roberta Labelle

9: **Clinical and Economic Evaluation of Diagnostic Technologies: The Cases of Impedance Plethysmography and Nuclear Magnetic Resonance** 167
Gordon Guyatt and Michael Drummond

10: **Clinical and Economic Evaluation of Therapeutic Technologies: The Case of Neonatal Intensive Care Programs** 199
Kathryn Bennett and David Feeny

11: **Policy Options for Health Care Technology** 225
Greg Stoddart and David Feeny

Members of the Institute 259

Institute Management 263

Publications Available 265

Foreword

National health care expenditures at present account for more than 8 per cent of the gross national product. The public sector expenditures of provincial governments on health (mainly health insurance programs) account for approximately 25 per cent of their budgets. Against this background several streams of public policy debate have focused attention on health technology policy. First, the high cost of health care has spawned concern over the efficiency of the health care system and strategies to control health care costs. Conflicts over funding between providers and government insurance programs have become increasingly frequent and serious. Second, conflicts among levels of government over responsibility for health policy and the funding of health care have become increasingly prominent. Discussion surrounding the enactment of the Canada Health Act (1984) further intensified interest.

A parallel set of issues, more specifically focused on health technology, have also emerged. One such issue is the contribution of technology to increases in health care costs. Operations such as coronary artery by-pass grafting, for which the direct cost of the procedure is approximately $9,500 (1983 dollars), and new diagnostic imaging systems such as nuclear magnetic resonance, with a capital and installation cost of $1 - 2 million, provide dramatic evidence of the potential for new technologies to affect health care costs. Moreover, low-cost-per-unit but frequently used procedures and tests may play a more subtle but greater role in contributing to health care costs.

Concern over health care technologies is not and should not be confined to their cost. Skepticism over the safety and effectiveness as well as the widespread and rapid adoption of new technologies before adequate evaluation are equally if not more important issues for public policy. Does the application of the latest technology really improve health? How can one know? Even if it does, are the improvements obtained thorugh the application of this technology great enough to offset the forgone opportunity of using the resources for a different technology? Is society making a wise use of its resources in adopting a particular technology, or are there better uses for the same resources elsewhere within the health care sector? Or would an expansion of social services, or other government programs, better serve the public interest?

These are difficult questions to answer. An approach to addressing them should contain two components. First, it should acknowledge the host of technical issues involved in the evaluation of health care technology. Second, it should make clear that the ultimate choice, though better informed by technical evaluation, cannot rest solely on such criteria. Instead, the choices must reflect the fundamental values and preferences of the members of our society. Technical evaluation can help assess the quantity of resources required for a particular end, but it cannot determine whether one investment of resources is more worthwhile than another.

This book is devoted to the first of the issues mentioned above – the development of a comprehensive set of guidelines for the clinical and economic evaluation of health care technologies. It sets out a rationale for their rigorous evaluation before their widespread adoption within the health care system. It establishes the kinds of evidence needed to conclude that a new technology is safe, effective, and efficient.

In addition to proposing that major new technologies be subjected to rigorous evaluation, often through randomized clinical trials, the book also proposes that a National Health Technology Assessment Council be formed with provincial government participation and a mandate to organize technology assessment evaluations and the dissemination of the results. Technology assessment in Canada has lacked a focus as well as adequate funding. Finally, a variety of policy options for major changes in the organization of the health care system are explored. A key to improving performance is that decision makers within the system be given clear incentives to use the evidence produced by such evaluations. Clinical budgeting and the promotion of competition among modes of health care delivery are among the promising policies that would give the providers of health care services the incentives to improve efficiency and adopt and effectively utilize new technologies.

This book should inform and stimulate much-needed public debate on the specific issues of health technology policy and on the broader issues of the organization and performance of the health care system.

Rod Dobell, President August, 1986

Avant-Propos

Les dépenses nationales consacrées aux soins de santé constituent à l'heure actuelle plus de 8 % du produit national brut du Canada. Quant aux dépenses engagées par les Provinces au chapitre de la santé (principalement pour les programmes d'assurance-maladie), elles représentent environ 25 % de leurs budgets. Dans un tel contexte, plusieurs courants de pensée dans le débat de politique ont attiré notre attention sur l'importance de la politique en matière de technologie de la santé. En premier lieu, le coût élevé des soins de santé a donné naissance à une certaine préoccupation quant à l'efficacité du système de santé et des stratégies visant à en réduire le coût. Ainsi, la gravité et la fréquence des conflits entre les fournisseurs de services et les régimes d'assurance gouvernementaux concernant le financement de ce système n'ont cessé de s'accroître. En second lieu, la responsabilité en matière de politique et de financement des soins de santé est une source grandissante de désaccords entre les paliers de gouvernement. Par ailleurs, le débat qui a entouré l'adoption de la Loi canadienne sur la santé (1984) a encore accru l'intérêt suscité par cette question.

Toute une série de questions parallèles, plus directement axées sur les techniques des soins de santé, sont également apparues. L'une de ces questions a trait à l'effet qu'exerce la technologie sur la hausse des coûts de la santé. Ainsi, des opérations chirurgicales comme le pontage aorto-coronarien par greffe, dont le coût direct est d'environ 9 500 dollars (de 1983), et les nouveaux systèmes d'imagerie médicale, tels que la résonance magnétique nucléaire, dont le prix et le coût d'installation atteignent 1-2 millions de dollars, sont une preuve frappante de l'influence potentielle des technologies nouvelles sur les coûts de la santé. Par ailleurs, il est très possible que les actes médicaux et les examens d'un coût unitaire peu élevé, mais qui sont d'un usage fréquent, influent de façon moins directe, mais aussi plus marquée sur la hausse des coûts de la santé.

Toutefois, le coût des techniques de soins n'est pas, et ne saurait être, le seul motif de préoccupation. En effet, le scepticisme à l'égard de la sécurité et de l'efficacité des technologies nouvelles, tout comme leur adoption rapide et généralisée avant un examen approprié devraient être des considérations tout aussi, sinon plus importantes en matière de politique d'État. L'utilisation de la technologie la plus récente est-elle réellement un facteur d'amélioration de la santé? Comment peut-on le savoir? Même si c'est le cas, les améliorations résultant de son application sont-elles assez notables pour compenser le manque à gagner résultant de la non-utilisation des mêmes ressources pour une autre technologie? La société fait-t-elle un bon usage de ses ressources en adoptant une technologie donnée, ou bien est-il possible d'utiliser ces mêmes ressources à meilleur escient dans un autre domaine du secteur de la santé? Ou encore, les intérêts du public seraient-ils mieux servis par une expansion des services sociaux ou la création d'autres programmes gouvernementaux?

Ce sont là des questions auxquelles il est difficile de répondre. Toute démarche en vue de trouver des solutions devrait comporter deux éléments. En premier lieu, elle devrait prendre en considération toute une foule de questions techniques qui entrent en jeu dans l'évaluation de la technologie des soins de santé. En second lieu, elle devrait établir clairement que le choix ultime, même s'il est mieux éclairé par l'évaluation technique, ne saurait reposer uniquement sur de tels critères. Au contraire, ce choix doit refléter les valeurs fondamentales et les préférences des membres de la société. L'évaluation technique peut nous aider à estimer quelle quantité de ressources est nécessaire à une fin particulière, mais elle ne saurait déterminer si un investissement de ces ressources est plus avantageux qu'un autre.

Ce livre est consacré à l'étude de la première de ces questions. Il formule en effet tout un ensemble de lignes directrices destinées à l'évaluation clinique et économique des technologies de la santé. Il explique les raisons de la nécessité de procéder à un examen rigoureux de ces technologies avant que leur usage ne se répande au sein du système de santé et définit les éléments requis pour parvenir à la conclusion qu'une nouvelle technologie est sans danger, efficace et efficiente.

En plus de proposer que les principales technologies nouvelles fassent l'objet d'une évaluation rigoureuse, principalement par le moyen d'essais cliniques aléatoires, ce livre suggère la création d'un Conseil national d'examen des technologies de la santé auquel participeraient les gouvernements provinciaux et qui aurait pour mandat d'organiser de tels examens et d'en diffuser les conclusions.

Jusqu'ici l'évaluation de la technologie au Canada a manqué d'une orientation claire aussi bien que d'un financement adéquat. Enfin, il explore un éventail d'options de politique destinées à opérer d'importants changements dans l'organisation du système de santé. L'un des éléments clés de l'amélioration de la performance est que les décideurs au sein de ce système doivent clairement être encouragés à faire usage des résultats des évaluations. L'établissement d'un budget au sein des services cliniques et l'incitation à la concurrence entre divers modes de prestation des services de soins sont parmi les mesures de politique les plus susceptibles d'encourager les fournisseurs de services de santé à améliorer le rendement ainsi qu'à adopter et à utiliser efficacement les nouvelles technologies.

Ce livre ne manquera pas de favoriser et d'éclairer un débat public, dont la nécessité s'impose, sur les questions de politique relatives à la technologie de la santé et sur les problèmes plus généraux de l'organisation et de la performance du système de santé.

Rod Dobell
Président

Août 1986

Preface

One newspaper headline today might announce new evidence of success in the treatment of moderate and severe high blood pressure (with anti-hypertensive drugs) or in the management of victims of heart attacks (with beta-blockers). Another, with equal veracity, might report that certain treatments currently offered by respected clinicians are doing more harm than good. (For example, clofibrate does reduce the level of cholesterol in the blood, but patients receiving this therapy also have elevated mortality rates.) Or, it could be reported that there is little evidence on how well a large number of frequently used tests and procedures really do perform (most uses of CT scanning are unevaluated; the role of exercise in preventing initial heart attack is not precisely known). In short, there are three co-existing categories of health care technologies: those that we know work, those that we know do not work but may still be in use, and those that may or may not work.

This "fact of life" poses great problems for the consumers of health care, patients and their families, the providers of care, public officials concerned with health care policy, and society as a whole. This book is about how to determine which health care technologies work, which ones do not, and under what conditions the effective ones can be expected to work—in short how to separate "the wheat from the chaff".

The book is divided into three sections. After the introduction in the first section, Chapters 2 and 3 describe the current situation. The second, Chapters 4 thorugh 10, develops guidelines for evaluating health care technology and offers examples of the application of those guidelines in real settings of clinical and economic evaluation. The third section, Chapter 11, examines public policy options for getting from where we are to where we would like to be. Although authorship of individual chapters has been assigned by identifying those primarily responsible for the preparation of each chapter, it should be understood that the entire group contributed to the whole manuscript through comment, advice, criticism, and active discussion.

The book is written with a broad audience in mind, including health care professionals, health policy makers, and most important, the consumers of health care services, the members of our society. The issues are complex, but the discussion that follows should help readers to understand better the quality of evidence needed to evaluate new technologies and offers them some guidance on how they might evaluate health care technologies themselves. The guidelines on health technology assessment presented in this book are designed to assist policy makers and practitioners in making the hard decisions about which new technologies to adopt and how to utilize them. Such guidelines are in no way a substitute for seasoned judgement, of course, but they may help to organize the information that must enter those judgements.

A key message of the book is the need for timely and rigorous clinical and economic evaluation of important new health care technologies before they are widely applied. Although in Canada we still have a long way to go to reach that goal, Canadians can take pride in being among those in the forefront of this movement. The authors and the Institute hope that this book will contribute to a focused, lively, and productive public policy debate on the issues involved in assessing health technology.

David Feeny
Hamilton, Ontario December, 1985

Acknowledgements

The preparation of this book was supported by the Institute for Research on Public Policy, the Canadian Medical Association, and McMaster University. A preliminary draft of the manuscript was provided to the Canadian Medical Association Task Force on the Allocation of Health Care Resources, chaired by Joan Watson, as background to their report, "Health: A Need for Redirection." We would like to thank Ms. Watson, Hilary Southall and the members of the Task Force, as well as Orvill Adams of the CMA, for their help and encouragement.

At McMaster a large number of colleagues (or former colleagues) assisted us by providing information and counsel and by commenting on draft manuscripts of various chapters. The short list includes Michael Boyle, Pat Caulfield, Amiram Gafni, Jack Hirsh, Russell Hull, Sargent Horwood, Wayne Keddy, John Laidlaw, Jonathan Lomas, David Sackett, Saroj Saigal, John Sinclair, Darrell Tomkins, George Torrance, and Chris Woodward. Though they should in no way be held responsible for the content of the book, the quality of their research in large part made this book possible.

Preliminary versions of portions of the book were presented at the Fifth Biennial Northeast Canadian/American Health Conference, Halifax, September 15-17, 1983, and at the Department of Epidemiology and Public Health, Yale University, February 14, 1984. Participants at those presentations provided helpful comments.

This project benefited at two stages from discussions with Clyde Behney, Denise Dougherty, Hellen Gelband, and other persons at the Health Program Branch, Office of Technology Assessment, U.S. Congress. Their help is gratefully acknowledged.

Discussions with Robert Evans and Morris Barer (University of British Columbia) and Theodore Marmor (Yale University) were helpful in formulating the policy options discussed in Chapter 11. Ralph Andreano, Jon Christianson, Frank Denton, Amiram Gafni, and Mel Kilman provided useful comments on an earlier version of Chapter 3.

We are also grateful to *Health Services Research, The New England Journal of Medicine,* and *Social Science & Medicine* for permission to reprint material from previously published papers.

Thoughtful and constructive criticism from anonymous referees and by the editorial staff of the IRPP also importantly contributed to the manuscript.

Finally, the staffs of the Department of Clinical Epidemiology and Biostatistics and of the Department of Economics of McMaster University cheerfully typed and retyped the manuscript and generally facilitated the project. Their help is also gratefully acknowledged.

David Feeny
Gordon Guyatt
Peter Tugwell

Contributors

Kathryn Bennett, M.Sc., is a research associate in the Department of Clinical Epidemiology and Biostatistics, McMaster University, and holds a graduate degree in medical sciences in design, measurement, and evaluation. She has collaborated on a number of research projects and has particular skills in critical appraisal, community effectiveness, and international public health.

Michael Drummond, D.Phil., was a visiting professor at McMaster on leave from the Health Services Management Centre, University of Birmingham, England, when much of the work on this book was completed. He has published two major works on economic evaluation in health care as well as numerous scientific papers. He has also served as a temporary adviser to the World Health Organization.

David Feeny, Ph.D., holds a joint appointment in the Departments of Economics and Clinical Epidemiology and Biostatistics at McMaster. He is currently engaged in research on the economic evaluation of prenatal diagnosis, rehabilitation programs, and the treatment of childhood cancer. His previous research includes a book and many articles on the economics of technological change and agricultural development.

Gordon Guyatt, M.D., M.Sc., is an internist and clinical epidemiologist in the Departments of Clinical Epidemiology and Biostatistics and Medicine at McMaster. His research program includes measurement of the effectiveness of diagnostic and therapeutic technologies and the quality of life, particularly in the cardio-respiratory area.

R. Brian Haynes, M.D., Ph.D., is an internist and clinical epidemiologist in the Departments of Clinical Epidemiology and Biostatistics and Medicine at McMaster. He has published widely on patient compliance, professional compliance, medical education, and the treatment of hypertension. He is also the chief epidemiologist on the international trial of extracranial-intracranial bypass surgery.

Roberta Labelle, M.A., is a lecturer in health economics in the Department of Clinical Epidemiology and Biostatistics at McMaster. Her current research includes investigations into health care utilization, health care financing, critical appraisal, health policy, criteria for doing economic evaluations in conjunction with clinical evaluations, and international public health.

Greg Stoddart, Ph.D., is the senior health economist in the Department of Clinical Epidemiology and Biostatistics at McMaster. He has performed economic evaluations of dialysis programs, home care programs, alcoholism therapy, screening programs, and diagnostic strategies for venous thrombosis. He has also written widely on major health policy issues including user fees, manpower planning, and health care financing. He has made methodological contributions in the areas of demand analysis and the inclusion of economic evaluation in clinical trials.

Peter Tugwell, M.D., M.Sc., is a physician and clinical epidemiologist, and chairman of the Department of Clinical Epidemiology and Biostatistics at McMaster. He has participated in the evaluation of many health care technologies and worked with major policy groups, including the Ontario Council of Health Task Force on High Technology. He has published widely on technology evaluation, the care of arthritic patients, medical education, compliance and international public health.

Summary

New health care technologies are frequently incorporated into the health care system without proper evaluation of their effectiveness and efficiency. Dissatisfaction with many technologies grows over time as performance fails to meet expectations, and a technology may fall into disuse as it is replaced by yet another new and incompletely evaluated technology.

Rigorous evaluation of the clinical effectiveness and economic efficiency of major technologies before their widespread adoption could improve performance within the health care system. This book is devoted to establishing guidelines for the standards of evidence for these evaluations. Although the technical issues are complex, unbiased clinical evaluation of therapeutic and diagnostic technologies must often rely on the method employed in establishing the effectiveness of new drugs – the randomized controlled clinical trial. In such studies patients are randomly assigned to groups receiving the alternative forms of therapy (or alternative tests) being investigated. These methods have proved to be more powerful than other methods in distinguishing effective from ineffective technologies. The results of such studies may then be used to promote the diffusion of effective new technologies while retarding the diffusion of ineffective or even harmful ones.

Because resources for health care are limited, it is important that only effective new technologies are adopted. Further, among effective methods relevant to the same condition, it is important that the most cost-effective methods are the ones employed so that the health care system produces the greatest possible improvement in health status for the population. Economic evaluation is therefore an integral part of the assessment of health technology. Such evaluations include an assessment of the resource use (cost) of the alternatives under consideration as well as an assessment of the outcomes. Because improvements in the quality of life and not just the extension of life are among the goals of providing care, special techniques for

assessing the effects of the quality of life are employed in assessing health technology.

This book develops an integrated and comprehensive set of guidelines for clinical and economic evaluation that allow for the determination of the safety, effectiveness, and efficiency of health care technologies. These guidelines are then applied to existing literature on several diagnostic and therapeutic technologies to illustrate their use.

The methods by which the results of such evaluations could be incorporated into the planning for introduction and utilization of new technologies are also described. The size and location of facilities needed to serve those who could benefit from the technology can be determined to promote its rational use.

Finally, the policy implications of technology assessment in the Canadian health system are addressed. One implication is the need for timely and rigorous evaluation of new technologies before their rapid adoption. The creation of a National Health Technology Assessment Council with provincial government participation and a mandate to produce and disseminate such evaluations would be a major step forward. The next step would be to promote a more rational health care delivery system in which providers have the incentive to use the results of evaluations. A number of policies such as clinical department budgeting (within hospitals) and the promotion of competition among modes of delivery (for the non-hospital practice setting) would align more closely the responsibilities for both financial and clinical decision making. A variety of other policy options are explored.

Abrégé

Il est fréquent que les nouvelles techniques de soins soient incorporées au système de santé sans que leur efficacité et leur rendement aient fait l'objet d'un examen adéquat. Progressivement, l'insatisfaction à l'égard de nombreuses techniques s'accroît du fait que leur performance ne répond pas aux attentes, et il arrive qu'une technologie soit abandonnée pour être remplacée par une autre technologie tout aussi nouvelle et insuffisamment évaluée.

L'examen rigoureux de l'efficacité clinique et du rendement économique des principales technologies de soins avant leur adoption généralisée pourrait améliorer la performance du système de santé. Ce livre a pour objet la formulation de lignes directrices permettant d'élaborer les normes requises pour l'établissement des preuves aux fins de ces évaluations. Malgré la complexité des problèmes techniques en jeu, une évaluation clinique impartiale des techniques thérapeutiques et diagnostiques doit souvent se baser sur la méthode utilisée pour éprouver l'efficacité de nouveaux médicaments, à savoir l'essai clinique aléatoire comparatif. Dans ce genre d'études, les patients sont répartis au hasard en groupes qui subissent les différentes formes de traitement (ou les différents examens) à l'étude. Ces méthodes se sont révélées supérieures à d'autres pour distinguer les technologies efficaces de celles qui ne le sont pas. On peut ensuite utiliser les résultats de ces études pour promouvoir la diffusion des nouvelles technologies efficaces, tout en faisant obstacle à celle des technologies inefficaces, voire dangeureuses.

Etant donné que les ressources affectées aux soins de santé sont limitées, il importe de n'adopter que les technologies efficaces. Qui plus est, parmi les méthodes efficaces qui s'appliquent à la même affection, il importe de n'employer que celles ayant le meilleur rapport coût-efficacité, de sorte que le système de santé produise la plus grande amélioration possible de l'état de santé de la population. L'évaluation économique fait donc partie intégrante de celle de la technologie de la santé. De telles évaluations comportent à la fois une estimation de l'utilisation des ressources (du coût) pour les différentes solutions à l'étude et une appréciation des résultats. Étant donné que l'amélioration de la qualité de la vie, et non pas uniquement sa prolongation, est l'un des objectifs du système de santé, le processus d'examen de la technologie de la santé doit employer des méthodes spéciales pour évaluer les effets qu'exerce cette technologie sur la qualité de la vie.

Cet ouvrage élabore un ensemble intégré et complet de lignes directrices destinées à l'évaluation clinique et économique des technologies de la santé et qui permettent d'en établir la sûreté, l'efficacité et le rendement. On applique ensuite ces lignes directrices, à travers la littérature du sujet, à plusieurs techniques diagnostiques et thérapeutiques afin d'illustrer leur emploi.

On y trouve également une description des méthodes grâce auxquelles il est possible d'incorporer les résultats de ces évaluations aux plans d'introduction et d'utilisation des nouvelles technologies. Il démontre aussi qu'il est possible de fixer la taille et l'emplacement des installations nécessaires pour répondre aux besoins des personnes susceptibles de tirer parti de ces technologies de manière à favoriser leur utilisation rationnelle.

Enfin, on y traite des conséquences politiques de l'évaluation de la technologie dans le secteur de la santé. L'une de ces conséquences est la nécessité d'une évaluation rigoureuse et opportune des nouvelles technologies avant que leur emploi n'ait le temps de se généraliser. La création au niveau fédéral d'un Conseil national d'examen des technologies de la santé auquel participeraient les gouvernements provinciaux et qui aurait pour mandat de produire et de diffuser de telles évaluations constituerait un grand pas en avant. L'étape suivante serait de promouvoir un système de prestation de services de soins plus rationnel, dans lequel les fournisseurs de services seraient encouragés à utiliser les résultats de ces évaluations. Un certain nombre de mesures de politique, telles que l'établissement d'un budget au niveau des services cliniques (dans les hôpitaux) et l'incitation à la concurrence entre divers modes de prestation des soins (parmi les praticiens en milieu non hospitalier), feraient coincider plus étroitement les responsabilités en matière de décisions tant cliniques que financières. Diverses autres options de politique sont également explorées.

Chapter 1:

Introduction: Health Care Technology

David Feeny

New health care technologies have generally been greeted with enthusiasm both by health care providers and the consumers of health care services, who have great faith in their effectiveness. Seldom, however, are new technologies fully evaluated before their widespread implementation. Enthusiasm for a technology frequently wanes as it fails in clinical experience to live up to advertised claims or hoped for benefits. The disappointments are often fueled by the high price of the new technology and augmented by the more general concern in our society over the rising cost of health care.

A variety of questions arise from this issue. How effective are the new technologies? Do they really work, and for what conditions and which patients? Do the benefits in improved health justify the costs? Is society obtaining a bargain by investing in these new technologies?

Questions like these arise daily within the health care system, often in the context of decisions about a specific program, service, or patient. Decisions must be and are being made continually. Are there methodological tools that could assist health planners, providers, and consumers in addressing these difficult questions? In the pages that follow we will argue that indeed there are a set of validated and powerful tools that can contribute to the thoughtful formulation of health technology policy. The use of such tools is not, however, a panacea; they can only help to inform decisions on the adoption and use of new technologies. Difficult social choices underlie such decisions and cannot and should not be avoided or obscured by the use of technical criteria.

Specifically, we will argue three points. First, major new and existing health care technologies should be rigorously evaluated to determine their clinical effectiveness before their widespread adoption. The first question to ask is, do they work? Second, evaluations should include economic and quality-of-life assessments. Are the improvements in health that this technology offers large enough, relative to the net cost, to favour its use over competing uses for the same resources? Third, widespread adoption of new technologies should be delayed pending the results of such rigorous clinical and economic evaluation.

Two major sets of policy implications follow. It will be necessary to create the capacity to identify major new technologies, conduct rigorous clinical and economic evaluations in a timely manner, and disseminate that information to key parties in the health care system – in short, to produce the kinds of information that policy makers and providers need to make more informed decisions. It will also be necessary to give policy makers and providers the incentive to use that information. A major characteristic of the Canadian health care system is that although a variety of policies promote cost containment, very few promote efficiency. Health technology policy therefore involves both the generation of information and general policy changes to promote a more rational health care system.

To develop these policy implications it will be important to examine the evidence on the role of health care technologies in improving health and in accounting for increases in the cost of health care. These issues are discussed in Chapter 2. Evidence on the typical patterns of technology diffusion in the health care system is taken up in Chapter 3. The intent in both chapters is to review the evidence and to set the stage for the policy discussions that follow. As will become clear, the existing evidence is far from perfect. It does, however, allow us to reach several conclusions: (1) new technologies do contribute to increases in the cost of health care, and (2) the widespread adoption of new technologies generally precedes the availability of hard evidence on their effectiveness and efficiency. Premature diffusion is altogether too common and can adversely affect both health and health care costs.

It is clear that new health care technologies become widely disseminated before they are rigorously evaluated; what is less clear is why. There are, however, a variety of insightful speculations on this issue. Many observers point to a technological style of practice reinforced by third-party payment schemes under which the patient directly bears only a fraction of the cost of the delivery of the service and providers are paid only for services rendered – the "technological imperative." A lack of standards for the critical evaluation of evidence in the training program for physicians may be another source of the predisposition toward the adoption of new technologies. Young physicians acquire much of their style of clinical practice during training at sophisticated teaching hospitals and by following the role models of respected senior clinicians. Even if their relatively intensive style of practice is appropriate for the case mix they encounter, because teaching hospitals are referral centres receiving more complicated, serious, or rare cases, such practice styles are unlikely to be appropriate for a standard community hospital or office practice. Financial incentives also affect adoption and utilization decisions (see discussion in Chapter 3). The faith that both physicians and patients have in technology, based in part on the notable advances of the 1940s and 1950s in particular, reinforces the tendency to assume that what is new is better. Finally, the practice of "defensive medicine" is thought to be important in a litigious society where malpractice suits are frequent.

The core of this book is the development and presentation of a comprehensive set of guidelines for health technology assessment (Chapters 4-7). Chapter 8 presents the methods by which the results of such clinical and economic evaluations can be incorporated into planning decisions about the provision and utilization of new technologies. The use of these guidelines is illustrated in case studies in Chapter 9 for diagnostic technologies and in Chapter 10 for therapeutic technologies.

The discussion focuses on the development of methodological standards for the evaluation of new technologies and the justification of the standards selected. It would not be correct to infer that these are the standards in use; in fact, a major reason for writing this book is that the current standards are only implicit and even then are not comprehensively or consistently applied.

The assessment guidelines are frequently discussed within the context of evaluating specific diagnostic and therapeutic technologies; but they can as readily be applied to the evaluation of drugs, rehabilitation programs, new technologies in dental care, new applications of computer technology in health care, and a wide variety of clinical, managerial, or organizational innovations. Thus, the methodology has broad application, though the specific discussions omit a number of potentially relevant topics.

The description and analysis of the current situation (Chapters 2 and 3) and lengthy development and application of the guidelines (Chapters 4 though 10) are followed by a distillation of the policy implications (Chapter 11): how to get from where we are to where we would like to be. The thrust of the book is how to alter the Canadian health care system so it will produce and use rigorous evidence on clinical effectiveness and economic efficiency while preserving the many accomplishments of the system in providing equity and access to health care. The problems are complex; the proposed changes are modest. Innovations in policy are likely to be incremental; change is likely to be gradual. Cumulatively, however, change can fundamentally alter the nature of the system. Canada is in a position today to be in the international vanguard in developing a rational and humane approach to health care technology by building on its earlier achievements in providing a high standard of health care to its citizens.

Chapter 2:

New Health Technologies: Their Effect on Health and the Cost of Health Care

David Feeny

Issues and Definitions

This book is devoted to the study of the issues surrounding the introduction and diffusion of innovations in health care. To pursue that goal the first task is to provide definitions of fundamental terms and concepts. The definition of technological innovation is a broad one. Innovation refers to the emergence of novelty in the knowledge of medical science, practice, or organization.[1] Health technology may be defined as "the set of techniques, drugs, equipment, and procedures used by health-care professionals in delivering medical care to individuals and the systems within which such care is delivered."[2] Thus, an innovation refers to a change in the set of technologies available for the provision of health care.

It is also important to lay a foundation for the discussion of health technology policy. First, it is necessary to discuss the health care system and its role in society from a broad perspective. Second, we need to explore the process of technological change and development itself.

Health care is important because people, the consumers of health care, value health. They value health both because it directly contributes to their satisfaction ("utility" in the jargon of the economist) and because it indirectly contributes to their satisfaction by increasing their productivity thereby enhancing their ability to earn the income needed to purchase goods and services (and leisure). Health care services are important, therefore, because those services, along with genetic endowment, environment, life style, and other factors, are important in maintaining health. Thus, the criteria for evaluating the effect of health care technologies focus on attempts to measure outcomes that are relevant to patients. The thrust of the argument is reflected in the criteria set forth in Chapters 4 through 7 for the assessment of health care technologies.

New technologies may be helpfully classified by use, form, and degree of embodiment. In classifying by use we can distinguish diagnostic, therapeutic (including curative, preventive, rehabilitative, and palliative therapies), supportive (facilities such as computer accounting), and organization technologies (for instance, health service or maintenance organizations versus fee-for-service). By form we have drugs, devices, and procedures.[3] A related distinction in economics is the degree of embodiment of the new technology, that is, the degree to which it is embodied ("contained") in a new drug or device. A new technology is disembodied if it is an idea or a procedure that is available without obtaining a new machine or drug in which the new technology is contained.

An example may make the distinction clearer. Computed tomography (CT) scanning is an example of an embodied technical change, even though the application of the machine to clinical practice involves a great deal of learning. A radiologist may have a thorough understanding of how and why a CT scanner works but will be unable to apply that technology to clinical practice unless the services of the machine itself are available, because the technology is embodied in the piece of capital equipment. The adoption of an embodied technological innovation requires both a commitment of financial resources, to purchase the object in which the new technology is embodied, and a commitment of human resources, to learn how to use the innovation.

Early ambulation following surgery is an example of a disembodied technological innovation. All that is required to adopt the innovation is to learn the indications and contraindications for its use and acquire additional understanding through actual practice. Adoption does not require the purchase of a piece of equipment becaue the innovation is an idea or procedure – because it is disembodied.

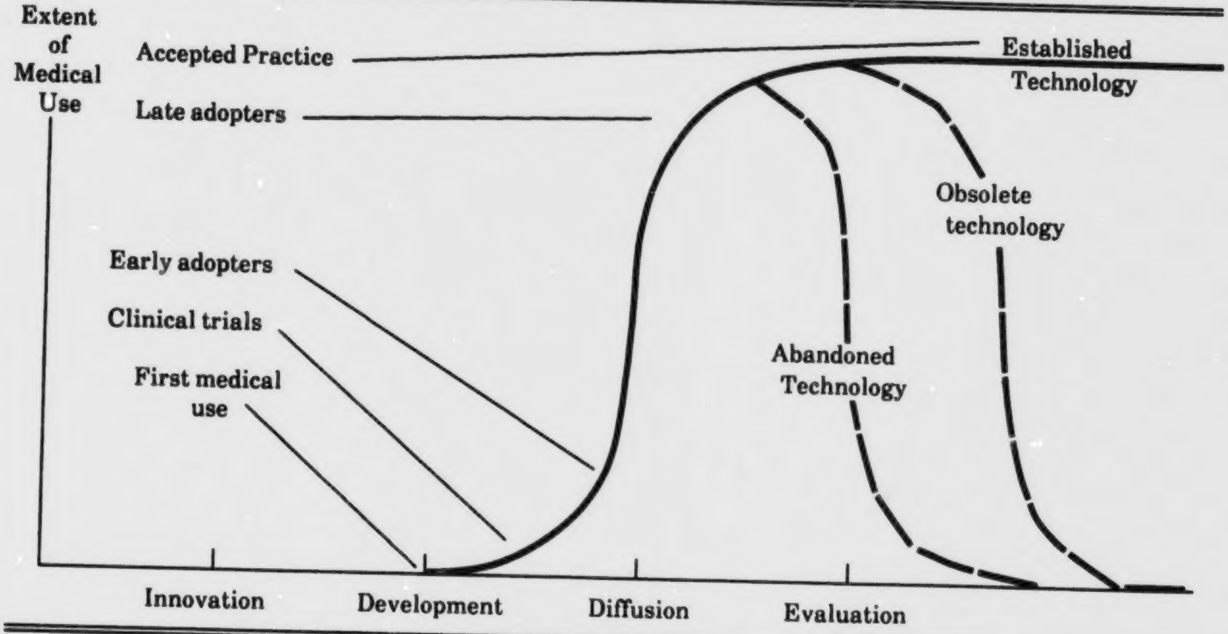
In practice, the distinction between embodied and disembodied technical change tends to blur. Most innovations are mixtures of both types of technical change. For instance, the neonatal intensive care unit, which will be examined in Chapter 10 as a case study, comprises a large number of procedures (many of which are examples of disembodied technical change) as well as a large number of specialized pieces of equipment (examples of embodied technical changes).

Stages in Technological Development

Four stages in the evolution of technical innovation in health care may be distinguished: innovation, development, diffusion, and evaluation (see Figure 2.1). Each will be discussed in turn.

FIGURE 2.1

Innovation, Development, and Diffusion of Medical Technologies


Extent of Medical Use
Accepted Practice
Late adopters
Established Technology
Obsolete technology
Early adopters
Clinical trials
First medical use
Abandoned Technology
Innovation
Development
Diffusion
Evaluation

Innovation is the emergence of novelty in the knowledge of medical science, practice, or organization. It is the first step in the process of the discovery of new medical technologies. It is often the result of research by either medical researchers or the suppliers of drugs and equipment to the health care sector.

There appear to be two implicit avenues through which the amount and direction of innovative activity in medicine are determined. In the first, private firms combine scientific knowledge and applied research to produce technological innovations, which are embodied in the products they hope to sell to the health care sector. Patent protection and the profits that result provide the incentives for the firms to invest in market studies to determine what kinds of innovations health care practitioners demand. The ability of the firm to capture much of the gain from innovation also provides the incentive for private firms to undertake the necessary development costs of introducing the innovation. As in other areas of innovation, the stock of scientific knowledge can have a profound effect on the cost and timing of new innovations (Rosenberg, 1974).

Because private innovators may not always be able to obtain sufficient returns, a substantial amount of medical and scientific research is publicly funded.[4] In this second implicit avenue, public granting agencies and the scientific community determine what kinds of research are needed and what kinds of innovations can be expected to have a high social pay-off. Because much of this research activity takes place at universities with medical colleges and at major health care and research institutes, researchers focus their attention on important clinical and public health problems. This interaction also assists in the further development, testing, and diffusion of medical technologies. Publicly funded research is particularly important for advances in the basic sciences and for disembodied innovations. Some research funding is designated for scientific disciplines such as biochemistry or genetics; much of it, however, is mission oriented and is focused on specific major diseases or health problems, such as cancer or heart disease.

In the innovation stage of technological development there are interactions among basic and applied research as well as among privately and publicly funded research and development activity. The development of the cardiac pacemaker provides an example. Fundamental advances in solid-state physics, developments in electrical engineering, the development of long-life batteries, advances in research on anatomy and cardiac surgery, and advances in materials science that resulted in improved electrodes, epoxy resins, and biocompatible materials all combined to produce the first successful human implant in 1960 (see Banta, Behney, and Willems, 1981, for more detail). Seemingly unrelated innovations in basic scientific knowledge in solid-state physics and the development of transistors were essential components along with the more specifically focused cardiac physiology and surgery studies that made pacemakers possible. Much of the basic scientific research was publicly funded; private manufacturers supported much of the refinement and applied development research.

Closely allied to the process of innovation is the process of development through which the new product or technique is refined and tested. Because the process is inherently iterative and cumulative, it is difficult to separate the period of innovation from the period of development.

Chalmers (1981, p. 325) distinguishes four phases in the development stage for drugs. By analogy these phases may be applied to other new health care technologies. Phase I occurs when the therapy is first administered to human patients. The goal of this phase is to ensure that the drug does not cause unacceptable toxicity. Animal experimentation and other investigations generally precede Phase I. Phase II comprises early uncontrolled trials on patients to provide a preliminary indication of potential effectiveness (if completely negative, the drug can be abandoned at this point). In Phase III large-scale, randomized trials of the new and old therapies are undertaken. Finally, in Phase IV the new therapy is used in practice but with some extra monitoring of outcomes.

A parallel system of classifying phases in the development stage is the distinction between the emerging technology that is still in the applied research stage and the new technology that has graduated to the clinical trials stage (Banta, Behney, and Willems, 1981, p. 9). New technologies then evolve into established ones that become part of standard practice.

The process by which a new technology becomes a standard practice is the process of diffusion or dissemination. It will be examined in Chapter 3. The behaviour both of organizations (hospitals) and of individual physicians is important in decisions about the adoption and utilization of new health care technologies. The key actors, however, are physicians who determine which services patients will receive, how frequently, and under what conditions.

After a technology becomes established it may receive further evaluation. The results of that evaluation (methods for evaluation are explored in detail in Chapters 4 through 7) may lead to the abandonment of the technology. More frequently the technology becomes obsolete in that it is replaced by a new one.

The stages of technological change—innovation, development, diffusion, and evaluation—are common to all sectors in the economy and do not apply exclusively to the health care sector. Two important themes commonly found in the history of technological change in other sectors are relevant to health technology as well. First, change is iterative and cumulative. Most innovations are the result of a myriad of minor advances that individually may appear to be unimportant but cumulatively have a major effect. A new technology is seldom developed in a single, major step. Modification and further technological development is an ongoing process. Second, technologies that are created initially for a specific use are often later adapted to a much broader spectrum of uses. The continual modification of new technologies and their spread to other uses makes the tasks of technology assessment and regulation more difficult than they would otherwise be. Indeed, a peculiarity of the health care sector is that the bulk of the evaluation of technology takes place after the diffusion process.[5] That problem is what this book is all about.

McKinlay (1981) has described the seven stages in the typical career of a medical innovation; these illustrate the need for mechanisms to ensure that rigorous evaluation takes place at an earlier stage:

1. promising report
2. professional and organizational adoption
3. public acceptance and state (third-party) endorsement
4. standard procedure and observational reports
5. randomized controlled trial (RCT)
6. professional denunciation
7. erosion and discreditation.

In the first stage, a "promising report" is issued on the efficacy of a medical innovation based on a few uncontrolled cases and perhaps followed by more extensive uncontrolled pilot studies. (The distinction between controlled and uncontrolled studies will be made clear in Chapter 5.) In the second stage, professional and organizational adoption, scattered support for the innovation increases to focused support by powerful interest groups and organizations. The innovation may be further legitimized through its inclusion in the curriculum of medical schools and through extensive unsystematic clinical experience.

In the third stage the innovation is accepted by the public and endorsed by third-party payers, usually without formal evaluation. The decision by insurers to reimburse is usually based on the innovation's prior acceptance by physicians rather than on rigorous evidence of its effectiveness.

In stage four the innovation becomes the standard practice and the subject of observational reports. Many of these case reports and retrospective studies are undertaken by persons or organizations interested in demonstrating the efficacy of the innovation. Unfortunately, these kinds of studies are also beset by serious methodological weaknesses (see Chapters 5 and 6).

In stage five, the randomized controlled trial (RCT) stage, rigorous evaluations of the clinical effectiveness of the innovation are finally undertaken. The conduct of the evaluation is complicated because the innovation has by now become standard practice, thus making the methodologically superior method of randomization more difficult. The results of the RCT are usually less favourable to the innovation than were the uncontrolled case reports. In addition, many innovations are never evaluated through the RCT and skip this career stage altogether.

McKinlay calls stage six the stage of professional denunciation during which the negative evidence that is often generated by the RCT is criticized. A double ethical and methodological standard seems to apply. Although it is unethical to withhold unevaluated treatment from patients because it is unevaluated (the difficulty in randomizing patients in stage five), it is ethical to treat patients with an unevaluated innovation without their informed consent. Positive case report evidence is apparently sufficient to underwrite the widespread diffusion of an innovation, but negative evidence from an RCT is subjected to intense methodological scrutiny. Such scrutiny is, of course, justifiable and proper; McKinlay is arguing only that each type of evidence should receive careful examination. RCTs that support current practice seem to receive more approval than those with negative results, which often meet a hostile and self-interested reception.

Time passes and the innovation moves into the seventh and final stage, the stage of erosion or discreditation. In a few cases a scandal leads to the abrupt abandonment of the innovation (such as thalidomide and diethylstilbesterol). More frequently it is replaced by a newer innovation.

This book describes how to move proper evaluation from stage five to stage two, that is, how to ensure that important health care innovations are subjected to rigorous evaluation before they become widely disseminated. At this point, however, it will be useful to review briefly some of the evidence on the effects of new health care technologies.

The Effect of Health Technology on Health: Historical Overview

New health care technologies affect society both by their effect on the health status of the population and by consuming resources that could have been used elsewhere. In some important cases, such as the advent of immunization for communicable diseases, new technologies have both improved health and reduced health care costs by reducing mortality and morbidity.

Today, new technologies most often improve health but increase the cost of health care in the process. In turn this means that fewer resources are available for other beneficial activities. In some cases new technologies have reduced health and increased cost, as in the use of clofibrate for the treatment of heart disease. Finally a new technology could worsen health and reduce cost. A primary function of health technology assessment is to identify technologies that improve health and reduce costs and to ensure that increased costs are justified by sufficiently improved health.

A rough estimate of the effect of biomedical research on health may be obtained by assessing how much of the reduction in mortality is attributable to economic, societal, environmental, provider characteristic, and technological variables. The twentieth century trends in Canadian mortality are presented in Table 2.1. Experience in the United States has been similar, as shown in Table 2.2.

All age groups enjoy higher life expectancy today than at the turn of the century, but the gains have been especially dramatic for infants and children. Between 1900 and 1975 the life expectancy at birth in the United States has increased by 25.2 years, whereas the average remaining years of life at age 60 have increased by only 4.5 years.

The mortality reductions reflect in part the late nineteenth- and twentieth-century scientific gains in the fields of bacteriology, chemistry, and biology. Fundamental understanding of the etiology of infectious and communicable diseases made a broad range of health-improving interventions possible, including the provision of sanitary water and food, insect and pest control, immunizations, antibiotics, and improvements in therapy and diagnosis. At the turn of the century diphtheria, pertussis (whooping cough), measles, typhoid, and scarlet fever each accounted for more child mortality than accidents; today, accidents are the leading cause of child mortality. As infectious and communicable diseases have declined as causes of death, cardiac disease and neoplasms (cancer) have increased.

TABLE 2.1

Life Expectancy by Sex at Selected Ages, Canada, 1871-1971

Age	1871[a]		1921[b]		1931		1961		1971	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
At Birth	n/a	n/a	n/a	n/a	60.0	62.1	68.4	74.2	69.3	76.4
20	47.9	47.3	49.1	49.2	49.1	49.8	51.5	56.7	51.7	58.2
40	33.4	33.6	32.2	33.0	32.0	33.0	33.0	37.5	33.2	39.0
60	18.2	18.2	16.6	17.1	16.3	17.2	16.7	19.9	17.0	21.4

[a] Includes Ontario, Quebec, Nova Scotia and New Brunswick.
[b] Includes all provinces except Quebec.

Source: Leacy, 1983; Series B65-74.

TABLE 2.2

Average Remaining Years of Life at Selected Ages, United States, 1900-1975

Age	1900	1930	1963	1975
At Birth	47.3	59.7	69.9	72.5
40	28.4	30.4	33.9	35.7
50	21.3	22.5	25.2	27.0
60	14.8	15.4	17.7	19.3

Source: Mushkin, 1979; p. 82

New health care technologies and biomedical research are now, however, responsible for all or even most of the decline in mortality. Higher incomes, better housing, better nutrition, safer work conditions, improved knowledge of health by the public, public health programs, increased provision of health care, and other factors have also contributed. Using U.S. data for 1900-1975 and 1930-1975, Mushkin (1979) has decomposed the mortality trend into the share of the decline accounted for by economic variables (measured by real per capita income), societal variables (measured by the unemployment rate as an index of stress), environmental variables (measured by the industrial accident rate), provider characteristics (a weighted average of the number of nurses and medical doctors per 100,000 people), and biomedical research (measured by the number of Ph.D.'s in biomedical science). Mushkin tried a number of alternative measures for most of her variables and found that the results were basically insensitive to the choice of measures. She found that from 1900 to 1975, biomedical research accounted for 30 to 40 per cent of the decline in mortality; from 1930 to 1975, for 20 to 30 per cent (Mushkin, 1979, p. 149). In a similar exercise for the years 1926-1974, biomedical research could account for roughly 39 per cent of the decline in morbidity, as measured by sick days among members of the armed forces during non-war periods (Mushkin, 1979, p. 252). Though biomedical research has had a significant effect on mortality and morbidity, it is not responsible for the bulk of the improvements in health status.[6]

As the relatively "easy" problems of infectious and communicable diseases have been "solved", the health care sector has become increasingly engaged in the treatment of conditions and diseases for which effective therapies are generally less available. Mushkin reports "burden-of-disease" estimates for 1900 and 1975 that reflect these trends. In 1900, indirect costs accounted for 90 per cent of the cost of disease (earnings forgone because of premature death or sickness); direct expenditures on health care accounted for only 10 per cent. By 1975, two-thirds of the costs were indirect and one-third were direct (Mushkin, 1979, p. 2).

Throughout the twentieth century, therefore, biomedical research in general and new health care technologies in particular have been important in reducing mortality and morbidity. More recently, gains have been less spectacular and have often come at the expense of greater health care costs.

The Effect of New Technologies on the Costs of Health Care

As discussed above some health care technologies have lowered the costs of health care through prevention or rapid and effective treatment (for example, the polio vaccine; surgery, radiation, and chemotherapy for Wilms' tumour; and antibiotic treatment of syphillis). In other cases expensive therapies have prolonged life and increased costs (such as therapies for lung cancer). To what extent have new technologies increased the costs of health care?

Because evidence on this question is conflicting, a precise answer is not possible. Several U.S. studies conclude that new technologies have contributed significantly to increased costs. Other studies show, however, that new technologies have played a much smaller role. Several issues arise in trying to determine the contribution of new technologies to the increases in costs. First, given the increase in real incomes and the spread of third-party payment schemes (health insurance), one would expect utilization of the health care system and thus health care expenditures to have increased. Second, costs rise as the prices of inputs (personnel, equipment, supplies, buildings) rise; the increase in costs should therefore be adjusted to reflect these price changes.

The largest number of studies are devoted to an examination of the role of technology in hospital cost increases. In a study of per diem hospital costs in the United States from 1951 to 1970, Waldman (1972) found that rises in the prices of inputs accounted for approximately one half of the overall increase and that increases in inputs per diem accounted for the other half. Klarman (1979) adjusts Waldman's figures to reflect changes in the age structure of the population and other factors and concludes that one-third of the increase was due to increases in inputs per diem. While recognizing that other factors are involved, some observers interpret the rise in inputs per diem to be a result of the introduction of new services, that is, the result of the adoption of new technologies (see discussions in Russell and Burke, 1975; Russell, 1979; Klarman, 1979; Davis, 1974; Cooper and Gaus, 1979; Altman and Blendon, 1979; and Banta, Behney, and Willems, 1981).

Looking at the rise in hospital operating costs, Banta, Behney, and Willems (1981, p. 89) report the results of a study by Redisch in which the more intense use of nine radiology and pathology services accounted for 40 per cent of the increase in costs from 1968 to 1971. The importance of ancillary services in accounting for cost increases is also highlighted by the work of Scitovsky and McCall (1976). They examined the trends in real inputs per patient for a number of common diseases or conditons in the actual practice of the Palo Alto Medical Clinic in California. For some diseases real costs fell; for some, real costs were roughly constant; and for others, real costs rose. In general, costs rose; the increases in laboratory tests and drugs per case more than compensated for the reductions in hospital days.

Using the same methodology in examining the trends in real costs per patient for ten common diagnoses at the University of California at San Francisco hospital, Showstack, Schroeder, and Matsumoto (1982) found that for 1972 to 1977 ther results only partially confirmed the earlier work of Scitovsky and McCall. For diagnoses in which standard practice technologies were used, there was little increase in cost. They also found that adjustments for case mix and the severity of the disease affected the results. The increased use of new technologies was therefore an important part of the cost increase in their hospital.[7]

Schroeder and Showstack (1979) also emphasize the importance of ancillary services in the rise in hospital operating costs. Fineberg (1979) points out that in 1950 approximately 100 laboratory tests were available to a practicing physician; by 1979, there were more than 600.

More recently a somewhat more sophisticated approach to examining the overall effect of new technologies on hospital costs has been employed. Like the methods discussed above, a fundamental problem arises in partioning the overall cost increase among the various factors responsible for the increase. In addition, because the quantitative measures available are often only loosely connected to the factor they purport to measure, interpretation of the results is complicated. In the newer method increases in total expenditures are attributed to the following changes: (1) in population, (2) in overall wages and prices, (3) in wages and prices in the health care sector greater than the general inflation, and (4) in service intensity (see U.S. Congress, 1984). As in Waldman (1972) and Klarman (1979) changes in service intensity (the real use of resource per episode of care), are interpreted as being attributable to technological change. In this approach, changes in wages and prices in the health care sector are similarly interpreted as reflecting the technical sophistication demanded by new technologies. These wages and prices may, of course, also reflect other factors, such as unionization of hospital workers. Because there is no readily quantifiable measure of technical change, indirect and crude measures are employed as substitutes.

Using methods such as these, Altman and Wallack (1979) found that in the United States from 1971 to 1976, one-third to one-half of the annual increase in daily hospital costs was associated with technical change. In a more recent U.S. study the Office of Technology Assessment estimated that increases in service intensity accounted for 24 per cent of the 93 per cent increase in per capita hospital costs between 1977 and 1982 (see U.S. Congress, 1984, for an excellent review of these and other estimates).

Although the figures range widely and the methodology is quite crude, one may conclude that new technologies are not the only reason health care costs have risen, but they are a major source of the rise in the costs of hospital care.

Mushkin (1979) has examined the role of new technologies in accounting for the rise in overall health care expenditures in the United States. She has decomposed the increase into the amounts attributable to population growth, rises in the prices of inputs, the growth in real per capita income, the spread of third-party payment systems, and the changing age structure of the population. One can assign the residual, the unexplained portion of the change in expenditures, to the use of new technologies. For 1930 to 1975, Mushkin found that the residual was negative. In more detailed analyses by disease she found negative residuals (where "technology" reduced the net costs of treatment) for infectious and parasitic diseases and positive residuals (where "technology" increased the net costs) for cancer, circulatory system and cerebrovascular diseases, mental disorders, and endocrine and metabolic diseases. These results are consistent with intuitive and casual observations of the progress of medical science. Effective preventive and therapeutic technologies are available for some infectious and communicable diseases and for diseases for which new technology has reduced mortality, morbidity, and health care costs while increasing the quality of life. For some intractable problems, new technologies have prolonged life while increasing health care costs. The assessment guidelines discussed in Chapters 4 through 7 address the issues of determining the quality and extent of prolonged life.

Using a methodology similar to that used by Mushkin, for the period 1947-1967, Fuchs found that technology increased health care costs, a positive residual of 0.6 of the 8.0 per cent annual increase. The period examined by Fuchs reflects to a lesser extent the earlier period covered by Mushkin during which there were major advances in immunizations and public health and sanitation. Thus it is not surprising that he found a positive residual (see also Mushkin, 1979; and Banta, Behney, and Willems, 1981).

The role of technology in the overall rise in U.S. health care expenditures has also been assessed more directly using the same methodology as was applied in the costs of hospital care studies. For 1977-1982, the Office of Technology Assessment found that the technology components are responsible for nearly 30 per cent of the rise in Medicare costs (U.S. Congress, 1984). Studies of the effect of specific technologies on the costs of health care are inconclusive; some show positive effects, others do not.

Overall, what do these studies tell us about the contribution of new technologies to the rise in the cost of health care? Before the answer to this question, three qualifications should be made. First, all of the methodologies are crude and have serious weaknesses. Second, these studies have focused on hospital costs and overall expenditures, leaving office practices and other non-hospital health care delivery underinvestigated. Third, little evidence is available on these issues for Canada; most of the evidence is for the United States.

New technologies do appear to have caused increases in health care costs, but the third-party payment system and other factors influencing the use of existing and new technologies are causes as well. A great deal of attention has been focused on "big ticket" items like CT scanners and nuclear magnetic resonance

(NMR) machines, but it is important to realize that the annual operating costs of even these very expensive machines are usually as much as 50 to 75 per cent of the purchase price (see Willems, Banta, Lukas, and Taylor, 1979; Sanders, 1979; and Kressley, 1981). Furthermore, technologies that have low per-unit costs but are frequently used, such as simple laboratory tests, have contributed as much, if not more, to the rise in health care costs (Schroeder and Showstack, 1979; Iglehart, 1977; Moloney and Rogers, 1979). The real question is, are the benefits of the applications of the new technologies worth the costs?

New and Emerging Technologies

Just as the diffusion of established technologies, such as intensive care units, coronary care units, cobalt therapy, CT scanners, coronary artery bypass surgery, various radiation and chemotherapies for cancer, and renal dialysis, have been associated with increases in the costs of health care, the technologies now emerging can be expected to play a similar role.[8] Biomedical engineering, the physical sciences, and the biological and biochemical sciences are generating a stream of increasingly sophisticated technologies.

Diagnostic innovations such as NMR (discussed in Chapter 9), digital radiography, fiber optics, maternal alpha-fetoprotein screening, positron emission tomography, automated karyotyping of chromosomes, chorionic villi sampling for prenatal diagnosis, light testing of skin, echocardiography, biochemical tests for Huntington's and other genetic diseases, new applications of ultrasound, retinal photography, cortical audiometry, new chemical analyzers, and serial pituitary hormone assays are being clinically applied or are being investigated and considered for application. A number of these innovations are based on recent advances in biochemistry and the application of computer assisted technologies to medicine.

New and emerging therapeutic technologies often build on improvements in diagnosis and diagnostic accuracy. New therapeutic technologies include intravascular balloon-catheter for vascular occlusion, electromyographic biofeedback, CT-linked radiation therapy, bone marrow transplantation, cerebral bypass surgery, new protocols in radiation and chemotherapy for a variety of cancers, therapeutic ultrasonics, fetal surgery to correct congenital malformations, organ transplants, new prosthetic cardiac valves and pacemakers, PUVA (psolaren plus ultra-violet light A) therapy for psoriasis, joint replacements, the development of ocular trauma units, lens implants, total parenteral nutrition, improvements in burn unit treatments, and the artificial pancreas. In addition to these diagnostic and therapeutic innovations, there are new drugs and screening tests and innovations in the delivery of health care, such as telemedicine.

In sum, there is no shortage of promising and important innovations in health care. If the problems of evaluating health care technology have been serious in the past, they are likely to become even more acute in the future. Clearly, resources must be devoted to evaluating the new technologies so that health care providers may choose the best among them and use them efficiently in the prevention, diagnosis, treatment, and palliation of diseases and other health problems.

Notes

1. This definition of innovation is a paraphrase of the one given in Hayami and Ruttan (1971, p. 54).

2. This definition is the one used by the United States Congress, Office of Technology Assessment (OTA), as quoted in Banta and Behney (1981, p. 447).

3. Amplification on the classificatory categories for medical technologies can be found in Banta, Behney, and Willems (1981).

4. Publicly funded research in this study includes research funded both by government and by charitable agencies. Economies of scale in research activities, positive spatial externalities due to technological borrowing (an advance developed in one jurisdiction is available to others), and negative externalities generated by contagious disease (a person affected by the disease may adversely affect another person by spreading the disease) provide further justification for the public sector to participate in medical research.

5. One could speculate that this is due to the informational asymmetries inherent in the health care sector. The ultimate consumer, the patient (the principal), seeks the advice and services of the health care professional (the agent) in part because that professional has differential expertise in the knowledge of health and health care. Because the consumer is not fully informed, he or she is not

in a position accurately to judge the performance of the physician. Thus, the physician may have a reduced incentive to provide the best care from the point of view of the patient. Medical ethics attempts to instill an ideology to overcome this problem.

6. See also Schneyer, Landefeld and Sandifer (1981). McKinlay and McKinlay (1977) argue that the introduction of specific medical measures and the expansion of medical services have been responsible for only a small portion of the modern decline in mortality. Increases in the standard of living and improved nutrition and hygiene are given larger roles in accounting for the decline in mortality. The results obtained by Mushkin (1979) partially corroborate these arguments; again it is important to distinguish between the contribution of biomedical research through improved public health and sanitation and through specific therapeutic interventions.

7. For a description of the effect of the addition of new facilities and technologies on costs and style of practice within Massachusetts General Hospital, see Sanders (1979); on the general point, see also Scitovsky (1979).

8. This list is not exhaustive or complete but is only meant to be suggestive.

References

Altman, Stuart H., and Blendon, Robert, eds. 1979. *Medical Technology: The Culprit Behind Health Care Costs?* DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Altman, Stuart H., and Wallack, Stanley S., 1979. "Is Medical Technology the Culprit Behind Rising Health Costs? The Case For and Against." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 24-38. DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Banta, H. David, and Behney, Clyde J., 1981. "Policy Formulation and Technology Assessment." *Milbank Memorial Fund Quarterly* 59(3): 445-479.

Banta, H. David; Behney, Clyde J.; Willems, Jane Sisk. 1981. *Toward Rational Technology in Medicine: Considerations for Health Policy*. New York: Springer Publishing Company, Inc.

Chalmers, Thomas C. 1981. "The Clinical Trial." *Milbank Memorial Fund Quarterly* 59(3): 324-339.

Cooper, Barbara S., and Gaus, Clifton R. 1979. "Controlling Health Technology." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Davis, Karen, 1974. "The Role of Technology, Demand and Labor Markets in the Determination of Hospital Costs." In *The Economics of Health and Medical Care*, ed. Mark Perlman, 283-301. London: MacMillan and Co. Ltd.

Fineberg, Harvey V. 1979. "Clinical Chemistries: The High Cost of Low-Cost Diagnostic Tests."In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 144-165. DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Hayami, Yujiro, and Ruttan, Vernon W. 1971. *Agricultural Development: An International Perspective*. Baltimore: Johns Hopkins University Press.

Iglehart, John K. 1977. "The Cost and Regulation of Medical Technology: Future Policy Directions." *Milbank Memorial Fund Quarterly* 55: 25-29.

Klarman, Herbert E. 1979. "Observations on Health Care Technology: Measurement, Analysis and Policy." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 273-291. DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Kressley, Donald M. 1981. "Diffusion of High Technology Medical Care and Cost Control – A Public Policy Dilemma." *Technology in Society* 3(3): 305-322.

Leacy, F.H., ed. 1983. *Historical Statistics of Canada 2nd ed.* Ottawa: Statistics Canada.

McKinlay, John B., and McKinlay, Sonja M. 1977. "The Questionable Contribution of Medical Measures to the Decline of Mortality in United States in the Twentieth Century." *Milbank Memorial Fund Quarterly* 55: 405-428.

McKinlay, John B. 1981. "From 'Promising Report' to 'Standard Procedure': Seven Stages in the Career of a Medical Innovation." *Milbank Memorial Fund Quarterly* 59(3): 374-411.

Moloney, Thomas W., and Rogers, David E. 1979. "Medical Technology – A Different View of the Contentious Debate over Costs." *New England Journal of Medicine* 301(26): 1413-1419.

Mushkin, Selma J. 1979. *Biomedical Research: Costs and Benefits.* Cambridge, Mass.: Ballinger.

Rosenberg, Nathan. 1974. "Science, Invention and Economic Growth." *Economic Journal* 84(333): 90-108.

Russell, Louise B. 1979. *Technology in Hospitals: Medical Advances and Their Diffusion.* Washington, D.C.: Brookings Institution.

Russell, Louise B., and Burke, Carol S., 1975. *Technological Diffusion in the Hospital Sector.* National Technical Information Service PB-245 642. Report prepared for the National Science Foundation, U.S. Department of Commerce.

Sanders, Charles A. 1979. "Technology and the Hospital." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 57-76. DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Schneyer, Solomon; Landefeld, Steven J.; Sandifer, Frank H. 1981. "Biomedical Research and Illness: 1909-1979." *Milbank Memorial Fund Quarterly* 59(1): 44-58.

Schroeder, Steven A., and Showstack, Jonathan A. 1979. "The Dynamics of Medical Technology Use: Analysis and Policy Options." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 179-212. DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Scitovsky, Anne A. 1979. "Changes in the Use of Ancillary Services for 'Common Illness.' " In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 39-56. DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Scitovsky, Anne A., and McCall, N. 1976. *Changes in the Cost of Treatment of Selected Illnesses, 1951-1964-1971*. DHEW Publication No. (HRA) 77-3161. Washington, D.C.: U.S. Government Printing Office.

Showstack, Jonathan A.; Schroeder, Steven A.; Matsumoto, Michael F. 1982. "Changes in the Use of Medical Technologies, 1972-1977: A Study of 10 Inpatient Diagnoses." *New England Journal of Medicine* 306(12): 706-712.

U.S. Congress, Office of Technology Assessment. 1984. *Medical Technology and the Costs of the Medicare Program*. OTA-H-227. Washington, D.C.: U.S. Government Printing Office.

Waldman, Saul. 1972. "The Effect of Changing Technology in Hospital Costs." U.S. Depoartment of Health, Education and Welfare, Social Security Administration, Office of Research and Statistics, Research and Statistics Note No. 4-1972 (February 28).

Willems, Jane Sisk; Banta, David H.; Lukas, Theresa A.; Taylor, Carl A. 1979. "The Computer Tomography (CT) Scanner." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 116-143. DHEW Publication No. (PHS) 79-3216. Washington, D.C.: U.S. Government Printing Office.

Chapter 3:

The Diffusion of New Health Care Technologies

David Feeny

Chapters 4 through 7 present guidelines for the clinical and economic evaluation of new technologies. Those chapters focus on how the system should operate; this chapter focuses on how the system does operate, that is, on how the use of new health care technologies spreads within the existing system.

An understanding of the diffusion process is important for several reasons. As Louise Russell (1979a) argues, "In order to evaluate the effects of regulation on technological innovation and diffusion . . . we need two benchmarks—a notion of how the industry ought to behave and a notion of how it would behave in the absence of regulation" (p. 26). Chapters 4 thorugh 7 deal with the former; this chapter is devoted to the latter. The goal is to ensure that effective and efficient new technologies diffuse rapidly while ineffective and inefficient ones do not diffuse at all. At present, most new technologies appear to diffuse rapidly, without a prior demonstration of their effectiveness.

The existing evidence and analysis of the diffusion of health care technologies will be briefly reviewed. The intent is not to provide definitive explanations for the trends observed. As will be evident there are significant gaps in the existing literature. Nonetheless, several useful conclusions emerge.

Issues in the Diffusion of Health Care Technology

A number of conceptually distinct issues are often confused in the literature on health technology and its diffusion. Given the complicated interrelationships among matters pertaining to technology, health care, and resource allocation, much of this confusion is natural. Discussions of health technology frequently consider five related issues: innovation, technological development, evaluation, diffusion, and efficiency. Chapter 2 examined innovation and technological development; Chapters 4 through 7 will cover evaluation and efficiency.

Two points from Chapter 2 require reiteration here. The first point is the distinction between embodied and disembodied technical change. An embodied technical change is contained in a new piece of equipment or a new drug. Learning how to use the new machine or drug is an important part of the diffusion process, but learning alone will not suffice because the technology is unavailable without the purchase of the machine or drug. In contrast, disembodied technical change refers primarily to a new procedure or technique and is thus available largely through learning without the purchase of new equipment or a new drug. The new technology is an idea; it is not primarily contained in a physical object.

The second point is that the development of new health care technologies is usually an iterative process comprising a series of small steps that individually may appear to be minor but cumulatively are important. Even while new technologies are undergoing improvements, modifications often continue to be made in older existing technologies (though generally at a slower rate). In general the services of highly skilled practitioners are required during the development phase (Newhouse, 1978; Rettig, 1979; Schneyer, Landefeld, and Sandifer, 1981). The development phase for embodied technical change is often privately funded by drug companies or equipment manufacturers; much of the development for disemsbodied technical change is conducted at teaching hospitals and is indirectly supported through the cost-reimbursement system of third-party insurers. As knowledge of the innovation accumulates through experience and investigation, its application becomes more routine, and a broader and less expert group of providers can use the innovation effectively. These ideas underlie the model of technology diffusion based on a theory of product and innovation cycles discussed below.

Research on the Diffusion of Health Care Technologies

Much research exists on diffusion in a wide variety of settings, including the medical sector (see Table 3.1). Numerous studies have made explicit or implicit use of the notion of constrained optimization (Approach 1, Table 3.1). These studies acknowledge the role of the goals of and constraints on the provider of health care services who decides when (or whether) to adopt a new technology. Investigators have then attempted to analyze the effects of the goals of the decision maker and the constraints that person faces on adoption decisions. The hospital's goals are usually assumed to be patient care (both the quality and quantity of care), teaching, and research. The hospital faces budgetary and other constraints in its pursuit of its goals. Seen in this light,

TABLE 3.1

Analytical Approaches to the Study of the Diffusion of New Health Care Technologies

	Approach	Major Assumption(s)	Evidence & Applications
1.	Constrained optimization	Adopts to further patient care, research and financial goals	Substantial evidence on the diffusion of hospital technologies that is consistent with the hypothesis
2.	Communication	Focuses on channels of communication of innovations and differences in access to information	Considerable application in tracing and explaining patterns of diffusion among new drugs
3.	Desperation-reaction model	Models treatment decisions; focuses on severity of disease as a major determinant of adoption	Application to diffusion of chemotherapy for leukemia
4.	Administration	Examines effects of centralization and other administrative variables on the organizational adoption of innovations	Case studies
5.	Innovation cycle model	Focuses on skills and learning as determinants of the rate of adaptation & adoption of innovations	Considerable application to the diffusion of agricultural innovations; little formal application in health care

Sources: See text

a hospital may decide to adopt a new technology because it will permit the hospital to improve patient care, recruit and retain high-quality staff, conduct research, or augment its revenues through billing for the new service.[1]

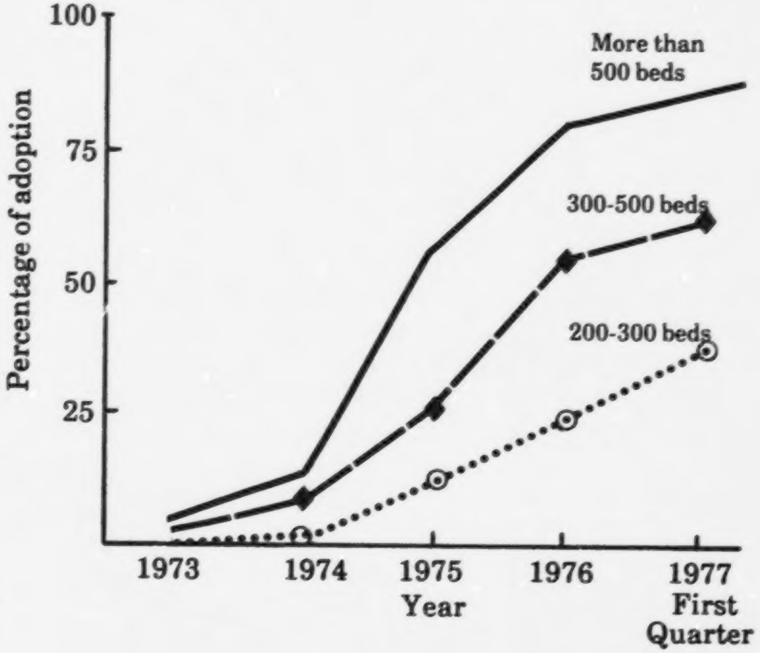
Studies conducted within this basic framework have tended to concentrate upon the adoption of new facilities by hospitals. Examples include the diffusion of intensive care units, post-operative recovery rooms, open-heart surgery facilities, diagnostic and therapeutic radiology, respiratory therapy, the electroencephalograph, computed tomography scanners, and dialysis (Baker, 1979; Banta, 1980; Rapoport, 1978; Russell and Burke, 1975; and Russell, 1976, 1977, 1979a, 1979b; Willems, Banta, Lukas, and Taylor, 1979). These studies conclude that the diffusion process may be represented by a logistic function,[2] that the relative advantage of adoption to the hospital in achieving its goals appears to be important in determining the rate of diffusion, that larger hospitals adopt new technologies first, and that third-party cost sharing speeds the rate of adoption. Hospitals affiliated with medical schools and hospitals with larger research programs tend to adopt new technologies sooner.[3] Adoption is also faster among hospitals in expanding markets, that is, in geographic areas where population is increasing (Banta, Behney, and Willems, 1981; p. 63).

In the context of office practice Schroeder and Showstack (1978) show that the existing fee structure promotes the adoption and use of tests and procedures at the expense of personal care by physicians. Evidence on the effects of the reimbursement system on hospital adoption of new technologies is found in Romeo, Wagner, and Lee (1984). They extend the work of Russell and others by examining the diffusion patterns for five relatively inexpensive capital-embodied hospital technologies in six states, three with prospective reimbursement and three without. Investigating the determinants of the probability, speed, and extent of adoption, the authors find that restrictive prospective reimbursement systems, such as that of New York State, tend to limit the extent of adoption but do not affect the probability and speed of adoption. For cost-saving new technologies, the New York program appears, in fact, to have increased the attractiveness of adoption. Financial incentives are important in affecting adoption and utilization decisions both in hospitals and among primary care physicians in a fee-for-service environment. Figure 3.1 illustrates the time path of the diffusion of a medical technology, CT scanners.

Another area of research on the diffusion of health care technology addresses the flow of communication and information

FIGURE 3.1

Adoption of CT Scanners, by Size of Hospitals
1973-1977



Source: Baker (1979), p. 157. Reprinted with the permission of *Social Science & Medicine*.

(Approach 2, Table 3.1). Coleman, Katz, and Menzel (1957, 1966) indicated that the early adoption of a new drug by physicians was related to the number of their subscriptions to medical journals, to their affiliations with medical institutions outside the community, to professional rather than patient orientation, and to the degree of their social integration in the medical community. Creditor and Garrett (1977) found that the time paths of publications on CT scanners and the adoption of CT scanners were very similar. For some cases of new findings in anesthesiology, Fineberg, Gabel, and Sosman (1978) found a similar pattern. Personal interactions also appear to be important in providing information on new innovations (Fineberg, Gabel, and Sosman, 1978; Manning and Denson, 1980; Rogers, 1971). Clearly, physicians and hospitals cannot adopt new technologies of which they are unaware.

Warner (1974a, 1975, 1977) studies the diffusion of new clinical therapies by modeling treatment decisions (Approach 3, Table 3.1). In his desperation model of treatment decisions he suggests that new innovations diffuse more rapidly (and in some cases prematurely) the greater the initial expectations of the efficacy of the new treatment, the more severe the case, the closer the medical problem is to one that has responded to similar treatment, the smaller the requirements for specialized staff and facilities, and the lower the cost of adoption. Warner also found in his study of the diffusion of leukemia chemotherapy that diffusion is affected by the physicians' learning from their own experience and from that of their colleagues.

A final major topic examined in the literature on the diffusion of health technology is the effect of the type of hospital administration on the diffusion process (Approach 4, Table 3.1). Gordon and Fisher (1975) present evidence that for high-risk innovations (innovations whose adoption might adversely affect the well-being of the hospital), the characteristics of the chief administrator were important and that the rate of adoption was positively related to the visibility of the consequences of adoption and negatively to the degree of centralization of the hospital administration. Kaluzny and Sprague (1974) found that innovations similar to ones that had already been adopted diffused more rapidly. Kimberly (1978) argues that the degree of centralization is unimportant for innovations that build on existing patterns but may be strongly related to adoption when new patterns of organization and behaviour are involved. Cosmopolitanism of organization members, training of the staff, and the administrator's attitudes toward change were important determinants of rates of adoption for low-risk innovations

(Becker, 1970; Gordon et al, 1974; Greer, 1977; Greer and Mohr, 1969).

Within economics there is a class of human capital models of technological adoption specifically designed to address diffusion issues. In these innovation cycle models (Approach 5, Table 3.1), producers with the highest skills are the first to adopt innovations because of their greater ability to recognize opportunities and solve adaptation problems (Kislev and Schchori-Bachrach, 1973; Evenson and Kislev, 1975; see also Vernon, 1966; and Welch, 1970). As information accumulates, less-skilled producers adopt the innovation and eventually, in a competitive industry, may drive the more-skilled producers out of the market. Under the existing cost-reimbursement system of third-party payers, consumers have little incentive to seek lower cost producers; higher-cost producers of medical services can, therefore, coexist with lower-cost, less-skilled producers. Schroeder and Showstack (1979) argue that the initial fee for a newly developed health care technology is generally high to allow providers to recoup their development costs and to compensate for initially low use. Charges are not usually lowered when utilization rates have risen and costs per unit have fallen, as they would be in a more competitive market (see also Bunker, Fowles, and Schaffarzick, 1982). Thus, in the health care setting of less than perfectly competitive markets, we may not observe the second part of the innovation cycle, the exit from the provision of a service by the highly skilled initial entrants. We would still expect, however, to observe the early entrance of the more-skilled producers. Skills are especially important in the development and early diffusion stages of innovation.

Empirical evidence from a wide variety of settings supports the hypothesis that highly-skilled producers tend to be early adopters (Kislev and Schchori-Bachrach, 1973; Evenson and Kislev, 1975). Eisenberg and Williams (1981) present evidence that the utilization of diagnostic technologies is inversely related to physician test scores on board examinations and that more knowledgeable physicians use tests more selectively, in short, that skills affect utilization rates. Finally, Kimberly (1978) presents evidence that the adoption of twelve innovations in the diagnosis, treatment, or prevention of respiratory disease by hospitals was positively related to the amount of research activity at the hospital, the number of paid outside speakers, and the reimbursement of expenses for physician attendance at professional meetings. Again, skills, as reflected in professional activities, research, and publications appear to affect the rate of adoption positively.

Conclusions

The brief review of the existing evidence and analysis points to several deficiencies in our current state of knowledge. First, very little of the evidence cited is from Canadian studies. Although impressionistic evidence indicates that the patterns observed in Canada are broadly similar to those in the United States, more information and analysis is needed within the Canadian context so that policy changes can be based upon a more accurate picture.

Second, the literature is heavily focused on the diffusion of embodied technologies (CT scanners, intensive care units, etc.) within the hospital sector. Very little evidence exists on the diffusion of disembodied technologies, such as surgical or laboratory techniques within the hospital sector; on the diffusion of new technologies (other than drugs) in the office practice of physicians; and on the effects of skills and learning on the nature and rate of the diffusion process in either delivery setting.

Third, studies that focused on the determinants of the rate of adoption usually failed to investigate the determinants of the extent of use (a partial exception is Romeo, Wagner, and Lee, 1984; see also Banta and McNeil, 1978; Banta, Behney, and Willems, 1981). Because most health care technologies are probably appropriate for some uses and inappropriate for others (often including many of their current applications), it is important to develop an understanding of the determinants of utilization, not just adoption. Potential remedies to the three major deficiencies in the literature on diffusion identified above will be briefly discussed in Chapter 11.

In spite of the deficiencies, several important conclusions about the nature of the diffusion process in health care emerge. New technologies are rapidly adopted because providers perceive adoption to be in their interests. Organizational behaviour may be involved, but key adoption and utilization decisions are in general made by or for physicians. Providers have multiple objectives in adopting new technologies, including financial considerations, prestige and collegial competition, and a sincere commitment to improving patient care and producing health benefits. Other factors that affect adoption and utilization decisions include the severity and urgency of the health care problem, the availability of alternatives, compatibility with the current style of practice, the prestige of the advocates for the new technology, the channels of communication, the climate of legal litigation, patient preferences, physician attitudes, and regulation (see Fineberg and Hiatt, 1979).

Several policy implications follow. Because providers are motivated by their desire to provide high-quality health care, the timely provision and communication of rigorous evidence on new technologies may affect adoption and utilization behaviour. Because providers also respond to financial and other incentives, policy makers may be able to affect adoption and utilization decisions by altering the incentives providers face. There are alternatives to direct regulation. These themes will be further developed in Chapter 11.

Notes

1. Russell and Burke (1975) suggest a hospital utility function in which the goals are patient care, teaching, and research. Davis (1972) surveys various models of hospital behaviour including recovery of costs, maximization of output, of output and quality, of utility, and of cash flow. Preliminary empirical evidence presented by Davis (1972) is mixed and does not clearly support any single one of these hypotheses. Data on the adoption of new specialized facilities by hospitals is consistent with the quantity-quality, utility, and cash flow maximization hypotheses (Davis, 1972). Lee (1971) suggests a model in which hospital administrators act to maximize utility by enhancing the status of their hospital and thus acquire inputs to maintain and improve the hospital's status in response to the behaviour of other hospitals; see also Becker (1970) and Rapoport (1978). Benjamini, Gafni, and Maital (1986) suggest that the hospital attempts to maximize the quantity and quality of services provided and that the latter is viewed as being closely related to the adoption of new technologies. They then model the interactions among hospitals in a given market area as a non-cooperative, non-zero sum game that frequently results in duplications of facilities. Newhouse (1970) uses a model in which the hospital maximizes the quality and quantity of patient care subject to budget and accreditation constraints. Feldstein (1971) develops a similar model. Russell and Burke (1975), Rapoport (1978), Reder (1965), Lee (1971), and Davis (1972) cite evidence that hospitals adopt new technologies to enhance their ability to retain and recruit high-quality staff; see also Robbins (1979), Schroeder and Showstack (1979), and Whitted (1981). Other investigators suggest that hospitals are run to maximize physicians' income; see Pauly and Redish (1973); Pauly (1978); Reder (1965); and Rosett (1974). Bredahl, Bryant, and Ruttan (1980) postulate a model in which the researcher maximizes a utility function

by allocating time between teaching and service, entrepreneurial activities (such as writing grant proposals), research, and leisure (as representative of non-work activities).

2. A number of investigators have found that the logistic function fits many, but not all, observed patterns of technological diffusion; see Banta, Behney, and Willems 1981; Coleman, Katz, and Menzel (1957, 1966); Rogers (1971); Romeo, Wagner, and Lee (1984); Rapoport (1978); Russell (1977); and Warner (1974b). The logistic function is:

$$P(t) = K / [1 + e^{-(a + bt)}]$$

where P(t) = the proportion of firms that have adopted the innovation by time (t); K is the proportion that will have adopted when the process of diffusion is complete; a is a constant defining the starting point of the process, and b is a constant measuring the rate of adoption.

3. It is common for new treatments and techniques to originate and be first tested in major teaching hospitals affiliated with medical colleges; see Baker (1979); Banta (1980); Bunker, Fowles, and Schaffarzick (1982); Gordon and Fisher (1975); Kressley (1981); Rettig (1979); Romeo, Wagner, and Lee (1984); Russell and Burke (1975); Russell (1979a, 1979b); Sanders (1979); Schroeder and Showstack (1979); and Warner (1974a).

References

Baker, Stephen R. 1979. "The Diffusion of High Technology Medical Innovation: The Computed Tomography Scanner Example." *Social Science & Medicine* 13D: 155-162.

Banta, H. David. 1980. "The Diffusion of the Computed Tomography (CT) Scanner in the United States." *International Journal of Health Services* 10: 251-269.

Banta, H. David; Behney, Clyde J.; Willems, Jane Sisk. 1981. *Toward Rational Technology in Medicine: Considerations for Health Policy*. New York: Springer.

Banta, H. David and McNeil, Barbara J. 1978. "Evaluation of the CAT Scanner and Other Diagnostic Technologies." *Health Care Management Review* 3: 7-19.

Becker, Marshall H. 1970. "Factors Affecting Diffusion of Innovations among Health Professionals." *American Journal of Public Health* 60: 294-304.

Benjamini, Y.; Gafni, A.; Maital, S. 1986. "The Diffusion of Medical Technology: A 'Prisoner's Dilemma' Trap?" *Socio-Economic Planning Sciences*. 20: 69-74.

Bredahl, Maury E.; Bryant, W. Keith; Ruttan, Vernon W. 1980. "Behavior and Productivity Implications of Institutional and Project Funding of Research." *American Journal of Agriculture Economics* 62: 371-383.

Bunker, John P.; Fowles, Jinnet; Schaffarzick, Ralph. 1982. "Evaluation of Medical-Technology and Strategies." *New England Journal of Medicine* 306: 620-624, 687-692.

Coleman, James; Katz, Elihu; Menzel, Herbert. 1957. "The Diffusion of an Innovation among Physicians." *Sociometry* 20: 253-279.

Coleman, James; Katz, Elihu; Menzel, Herbert. 1966. *Medical Innovation: A Diffusion Study*. Indianapolis: Bobbs-Merrill.

Creditor, Morton C., and Garrett, Julie Beetle. 1977. "The Information Base for Diffusion of Technology: Computed Tomography Scanning." *New England Journal of Medicine* 297: 49-52.

Davis, Karen. 1972. "Economic Theories of Behavior in Non-profit Private Hospitals." *Economics and Business Bulletin* 24: 1-13.

Eisenberg, John M., and Williams, Sankey V. 1981. "Cost Containment and Changing Physicians' Practice Behavior: Can the Fox Learn to Guard the Chicken Coop?" *Journal of the American Medical Association* 246: 2195-2201.

Evenson, Robert E., and Kislev, Yoav. 1975. *Agricultural Research and Productivity*. New Haven: Yale University Press.

Feldstein, Martin S. 1971. "Hospital Cost Inflation: A Study of Nonprofit Price Dynamics." *American Economic Review* 61: 853-872.

Fineberg, Harvey V.; Gabel, Ronald A.; Sosman, Martha B. 1978. "Acquisition and Application of New Medical Knowledge by Anesthesiologists." *Anesthesiology* 48: 430-436.

Fineberg, Harvey V., and Hiatt, Howard H. 1979. "Evaluation of Medical Practices: The Case for Technology Assessment." *New England Journal of Medicine* 301: 1086-1091.

Gordon, Gerald, et al. 1974. "Organizational Structure, Environmental Diversity, and Hospital Adoption of Medical Innovations." In *Innovation in Health Care Organizations: An Issue in Organizational Change*, eds. Arnold D. Kaluzny, John J. Gentry, and James E. Veney, 53-75. Department of Health Administration, School of Public Health, University of North Carolina at Chapel Hill, Monograph Series No. 4.

Gordon, Gerald, and Fisher, G. Lawrence. 1975. *The Diffusion of Medical Technology*. Cambridge, Mass.: Ballinger.

Greer, Ann Lennarson. 1977. "Advances in the Study of Diffusion of Innovation in Health Care Organizations." *Milbank Memorial Fund Quarterly* 55: 505-532.

Greer, Ann Lennarson, and Zakhar, Arlene A. 1979. "Patient Leverage Theory Proves to Be False." *Hospitals* 53: 98-106.

Kaluzny, Arnold D. 1974. "Innovation in Health Services: Theoretical Framework and Review of Research." *Health Services Research* 9: 101-120.

Kaluzny, Arnold D., and Sprague, Jane B. 1974. "Innovations in Health and Welfare Organizations: A Review and Critique of Current Theory and Research." In *Innovation in Health Care Organizations: An Issue in Organizational Change*, eds. Arnold D. Kaluzny, John T. Gentry, and James E. Veney, 1-14. Department of Health Administration, School of Public Health, University of North Carolina at Chapel Hill, Monograph Series No. 4.

Kaluzny, Arnold D.; Veney, James E.; Gentry, John T. 1974. "Innovations of Health Services: A Comparative Study of Hospitals and Health Departments." In *Innovation in Health Care Organizations: An Issue in Organizational Change*, eds. Arnold D. Kaluzny, John T. Gentry, and James E. Veney (77-110). Department of Health Administration, School of Public Health, University of North Carolina at Chapel Hill, Monograph Series No. 4.

Kimberly, John R. 1978. "Hospital Adoption of Innovation: The Role of Integration into External Informational Environments." *Journal of Health and Social Behavior* 19: 361-373.

Kislev, Yoav, and Schchori-Bachrach, Nina. 1973. "The Process of an Innovation Cycle." *American Journal of Agricultural Economics* 55: 28-37.

Kressley, Konrad M. 1981. "Diffusion of High Technology Medical Care and Cost Control – A Public Policy Dilemma." *Technology in Society* 3: 305-322.

Lee, Maw Lin. 1971. "A Conspicuous Production Theory of Hospital Behavior." *Southern Economic Journal* 38: 48-58.

Manning, Phil R., and Denson, Teri A. 1980. "How Internists Learned about Cimetidine." *Annals of Internal Medicine* 92: 690-692.

Mohr, Lawrence B. 1969. "Determinants of Innovation in Organization." *American Political Science Review* 63: 111-126.

Newhouse, Joseph P. 1970. "Towards a Theory of Nonprofit Institutions: An Economic Model of a Hospital." *American Economic Review* 60: 64-74.

Newhouse, Joseph P. 1978. *The Economics of Medical Care: A Policy Perspective*. Reading, Mass.: Addison-Wesley.

Pauly, Mark V. 1978. "Medical Staff Characteristics and Hospital Costs." *Journal of Human Resources* 13: 77-111.

Pauly, Mark V., and Redish, Michael. 1973. "The Not-for-Profit Hospital as a Physicians' Cooperative." *American Economic Review* 63: 87-99.

Rapoport, John. 1978. "Diffusion of Technological Innovation among Nonprofit Firms: A Case Study of Radioisotopes in U.S. Hospitals." *Journal of Economics and Business* 30:108-118.

Reder, M.W. 1965. "Some Problems in the Economics of Hospitals." *American Economic Review* 55: 472-480.

Rettig, Richard A. 1979. "End-Stage Renal Disease and the 'Costs' of Medical Technology." In *Medical Technology: The Culprits Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 88-115. Washington, D.C.: U.S. Government Printing Office.

Robbins, Frederick C. 1979. "Assessing the Consequences of Biomedical Research." In *Medical Technology: The Culprits Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 166-177. Washington, D.C.: U.S. Government Printing Office.

Rogers, Everett M. 1971. *Communication of Innovations: A Cross-Cultural Approach, 2nd ed.* New York: Free Press.

Romeo, Anthony A.; Wagner, Judith L.; Lee, Robert H. 1984. "Prospective Reimbursement and the Diffusion of New Technologies in Hospitals." *Journal of Health Economics* 3: 1-24.

Rosett, Richard N. 1974. "Proprietary Hospitals in the United States." In *The Economics of Health and Medical Care,* ed. Mark Perlman, 57-65. London: MacMillan and Co., Ltd.

Russell, Louise B. 1976. "The Diffusion of New Hospital Technologies in the United States." *International Journal of Health Services* 6: 557-580.

Russell, Louise B. 1977. "The Diffusion of Hospital Technologies: Some Econometric Evidence." *Journal of Human Resources* 12: 482-502.

Russell, Louise B. 1979a. "Regulating the Diffusion of Hospital Technologies." *Law and Contemporary Problems* 43: 26-42.

Russell, Louise B. 1979b. *Technology in Hospitals: Medical Advances and Their Diffusion*. Washington, D.C.: Brookings Institution.

Russell, Louise B., and Burke, Carol S. 1975. *Technological Diffusions in the Hospital Sector*. National Technical Information Service, PB-245 642. Report prepared for the National Science Foundation, U.S. Department of Commerce.

Sanders, Charles A. 1979. "Technology and the Hospital." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon 57-76. Washington, D.C.: U.S. Government Printing Office.

Schneyer, Solomon; Landefeld, J. Steven; Sandifer, Frank H. 1981. "Biomedical Research and Illness: 1909-1979." *Milbank Memorial Fund Quarterly* 59: 44-58.

Schroeder, Steven A., and Showstack, Jonathan A. 1978. "Financial Incentives to Perform Medical Procedures and Laboratory Tests: Illustrative Models of Office Practice." *Medical Care* 16: 289-298.

Schroeder, Steven A.; and Showstack, Jonathan A. 1979. "The Dynamics of Medical Technology Use: Analysis and Policy Options." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon 178-212. Washington, D.C.: U.S. Government Printing Office.

Vernon, Raymond. 1966. "International Investment and International Trade in the Product Cycle." *Quarterly Journal of Economics:* 190-207; reprinted in *The Economics of Technological Change*, ed. Nathan Rosenberg, 440-460. Baltimore: Penguin Books.

Warner, Kenneth Edgar. 1974a. "The Diffusion of Leukemia Chemotherapy: A Study in the Nonmarket Economics of Medical Care." Ph.D. diss., Yale University.

Warner, Kenneth Edgar. 1974b. "The Need for Some Innovative Concepts of Innovation: An Examination of Research on the Diffusion of Innovations." *Policy Sciences* 5: 433-451.

Warner, Kenneth Edgar. 1975. "A 'Desperation-Reaction' Model of Medical Diffusion." *Health Services Research* 10: 369-383.

Warner, Kenneth Edgar. 1977. "Treatment Decision Making in Catastrophic Illness." *Medical Care* 15: 19-33.

Welch, Finis. 1970. "Education in Production." *Journal of Political Economy* 78: 35-39.

Whitted, Gary S. 1981. "Medical Technology Diffusion and Its Effects on the Modern Hospital." *Health Care Management Review* 6: 45-54.

Willems, Jane Sisk; Banta, H. David; Lukas, Theresa A.; Taylor, Carl A. 1979. "The Computed Tomography (CT) Scanner." In *Medical Technology: The Culprit Behind Health Care Costs?* eds. Stuart H. Altman and Robert Blendon, 116-143. Washington, D.C.: U.S. Government Printing Office.

Chapter 4:

A Framework for the Evaluation of Technology: The Technology Assessment Iterative Loop

---

Peter Tugwell, Kathryn Bennett, David Feeny,
Gordon Guyatt & R. Brian Haynes

The increasing interest in evaluation of health care technologies has resulted in the development of new systematic approaches to this complex task. The overall technology assessment framework adopted for this book, the Technology Assessment Iterative Loop (TAIL) was developed to provide a systematic approach to the assembly of those specific subsets of health information necessary for rational decision making on the diffusion of health care technology (see Figure 4.1). This framework subdivides the spectrum of health information into sub-groups that constitute a logical progression from quantifying the burden of disease, through use of technologies to diagnose the cause of disease, to validating intervention that prevents or ameliorates disease, to selecting the application and diffusion of these interventions, and coming full circle, to determining whether the burden of disease has been reduced. Each of the seven steps in the Technology Assessment Iterative Loop poses a different type of research or evaluation question, and if new information is required, calls for a specific set of methods.

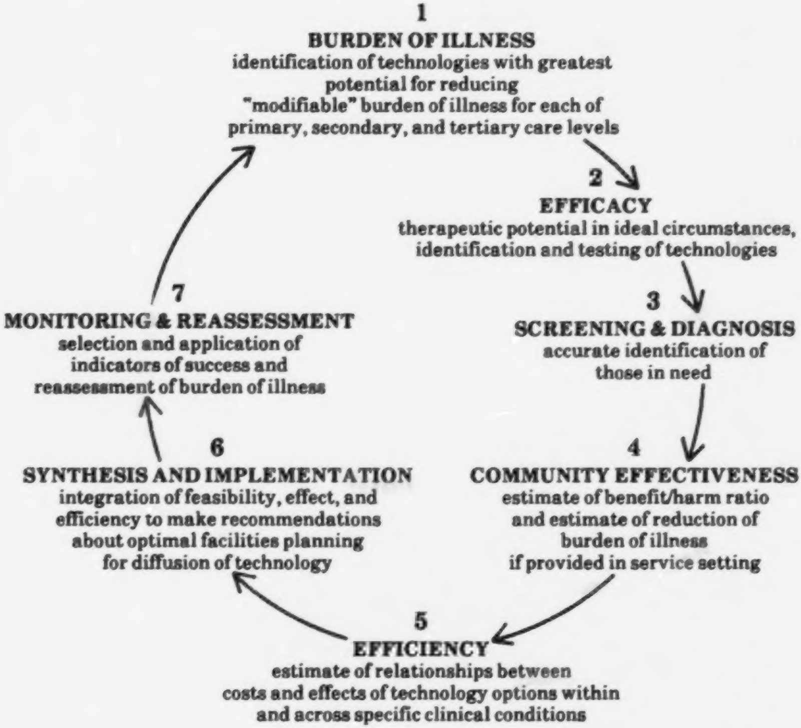
The "loop" format emphasizes the importance of monitoring after implementing a health intervention to determine whether the planned reduction in the burden of illness is achieved. Its practical guide to assessment methods is useful both to the "consumers" of research, health professionals and policy makers who decide whether to *apply* the results of research to health care decisions, and to the "producers" of research, those who plan and implement health care technology research.

Step 1: Burden of Illness

This first step determines current levels of morbidity (symptoms: physical, emotional, and social disability) and mortality (death rates, healthy years of life lost) using data sources, such as national and provincial data systems, and special sources, such as the *Canada Health Survey* (Canada, Health and Welfare Canada, 1981). The information on burden of illness needs to be checked

FIGURE 4.1

The Technology Assessment Iterative Loop


1
BURDEN OF ILLNESS
identification of technologies with greatest potential for reducing "modifiable" burden of illness for each of primary, secondary, and tertiary care levels
2
EFFICACY
therapeutic potential in ideal circumstances, identification and testing of technologies
3
SCREENING & DIAGNOSIS
accurate identification of those in need
4
COMMUNITY EFFECTIVENESS
estimate of benefit/harm ratio and estimate of reduction of burden of illness if provided in service setting
5
EFFICIENCY
estimate of relationships between costs and effects of technology options within and across specific clinical conditions
6
SYNTHESIS AND IMPLEMENTATION
integration of feasibility, effect, and efficiency to make recommendations about optimal facilities planning for diffusion of technology
7
MONITORING & REASSESSMENT
selection and application of indicators of success and reassessment of burden of illness

for its quality because we are proposing that it form a major criterion for health policy decisions and allocation of funding for research and development. The accuracy of this information on burden of illness is also crucial as a baseline for assessing the effect of new technologies. We propose three guidelines for assessing the quality of this information.

First, is the attribute that is selected for the measurement relevant? For example, mortality rates tell us nothing about diseases such as arthritis, whose major contribution to burden of illness is through symptoms and disability.

Second, are the measurement methods accurate? Measurement of the health status attribute (sometimes called a health status indicator) must be accurate. Cause of death is often inaccurate, and most components of morbidity require specially designed surveys. Caution should be exercised when information on burden of illness is obtained from surrogate measures of utilization or supply; utilization and supply data have been shown to be at variance with good survey data in some situations (Kohn and White, 1978). A good example is the enormous variation in technology utilization (such as surgical rates for hysterectomies and cholecystectomies) in populations with similar morbidity rates (Roos and Roos, 1982). Other sources of measurement error need to be avoided. In blood pressure studies, for example, conducting only one blood pressure assessment per person may result in an overestimate of the prevalence of high blood pressure; many patients who have elevated blood pressure on a single reading will have normal blood pressure on subsequent readings. When good evidence is not available, it is particularly important to check multiple sources of data.

Third, are the results easy to interpret and apply? Summary statistics (with ranges, means, and confidence limits) without mathematical transformation or sophisticated statistics are often sufficient. When morbidity and mortality are combined in the assessment of the effect of diseases or decrease in the quality-adjusted life expectancy, the assumptions and relative weights need to be stated (as they are in the important economic analysis of resource allocation in the management of hypertension by Stason and Weinstein, 1977).

Modifiable and Unmodifiable Burden of Illness

The usefulness of applied research for developing health care technology varies according to the breadth and depth of the available basic knowledge of specific health problems and the discovery of technologies with potential for prevention, cure, or palliation. Accordingly, burden of illness can be described as "unmodifiable" and "modifiable."

Unmodifiable burden of illness includes disability, symptoms, and mortality for which no effective interventional medical technology exists, such as disseminated cancer, multiple sclerosis, and Alzheimers disease. Resource allocation here should focus on research into etiology, prevention, and cure and not only on the care of current victims. Research funds should be applied to basic research in physics, chemistry, and biology, to overcoming fundamental technical problems, and to defining potential applications of the new technology to health care.

Modifiable burden of illness includes disability, symptoms, and mortality for which health care technology exists that *has* the potential to provide accurate diagnosis and efficacious prevention, cure, or palliation. Resource allocation here should be provided for the demonstration of effectiveness of technologies that have the potential to improve health care, and then for the development and diffusion of the optimal applications of these technologies for each of primary, secondary, and tertiary care levels, as described in the subsequent steps of the Technology Assessment Iterative Loop.

Step 2: Efficacy (Therapeutic Potential)

Health technology assessment should be anchored to the ability of the new technology in question to improve health directly or indirectly. Because the usefulness of technologies directed at screening and diagnosis ultimately depends upon their influence on subsequent therapeutic interventions, we will establish criteria for the efficacy of health interventions that employ health care technology before turning our attention to diagnostic technologies.

Exploratory studies (also called efficacy studies) ask the question, can it work? That is, they ask whether a therapeutic technology does more good than harm to patients who are diagnosed correctly, who are appropriately cared for, and who fully comply with recommendations for treatment. Exploratory studies provide information on the technology's potential, under ideal conditions, to ameliorate or to eliminate the effects of a disease or illness and about the potential toxicity or harm that the treatment itself might cause.

The formal testing of efficacy is unjustified unless the development of the technology has advanced to the point where the potential benefits and toxicity can be estimated with sufficient accuracy and unless the results of the initial studies are sufficiently promising for patients ethically to be exposed to the new technology.

The developmental stage of a new technology involves considerable exploratory research in basic science laboratories. The development of a new member of the most ubiquitous class of therapeutic technologies, drug therapy, often begins with the manipulation of an agent's molecular structure. What follows is the establishment of the drug's effect, and its safety, in laboratory animals, after which the agent is first administered to human beings. When a new surgical procedure is developed, investigators may face multiple challenges in the course of animal experimentation. The successful application of coronary artery bypass surgery, for example, involved the development of a satisfactory method of stopping the heart, maintaining blood flow to the tissues while the heart was not pumping, and overcoming the technical difficulties of the surgery itself.

On the basis of both theoretical constructs and fulfilment of technical specifications the potential applications of the new technology can be explored. Having established that cimetidine effectively inhibits the production of acid from the stomach, one might anticipate its being effective in the treatment of duodenal ulcer, gastric ulcer, and possibly gastrointestinal bleeding secondary to these problems. Coronary artery bypass surgery may alleviate symptoms of angina suffered by those with atherosclerotic heart disease, and might even prolong life, but these effects may differ depending on the severity of the underlying disease, the number and location of involved vessels, and the level of underlying cardiac function.

Because exploratory (efficacy) studies form the major basis of the estimate of benefit from a new technology, careful attention needs to be applied to the design issues of such studies. The important issues to be considered in implementation of efficacy are addressed in detail in Chapter 5.

Step 3: Diagnosis and Screening

Accurate detection or screening of patients is usually important for correctly identifying those individuals who will (and those who will not) benefit from efficacious interventional technologies and for estimation of prognosis. Technologies for diagnosis and screening probably are the fastest growing area in health care technology, with such innovations as computed tomography (CT), nuclear magnetic resonance (NMR), and positive emission tomography.

Many screening and diagnostic technologies identify "unmodifiable" health needs; this not only is an unnecessary expense but can actively contribute to the negative health consequences of the "labelling" that results. A study of careful follow-up of steel workers screened for hypertension revealed that absenteeism rose among previously unaware hypertensives labelled as a result of the screening process (Haynes et al., 1978); this led the authors and three Canadian task forces on hypertension to recommend that detection of hypertension should be carried out only in settings where adequate therapy (which has been proven to be efficacious) and long-term follow-up are ensured.

This concern for the necessity of linking diagnostic technology to efficacious therapy and for improving outcomes led to the following hierarchy for studies of diagnostic and screening technologies.

1. Technological capability. The performance specifications are made in a laboratory setting.
2. Diagnostic accuracy. The technology provides information that allows the health worker to make a more accurate assessment about the presence and severity of disease.
3. Effect on the health care provider. The technology allows health workers to be more confident of their diagnoses and thereby decreases anxiety and increases comfort.
4. Therapeutic effect. The therapeutic decisions made by the health care provider are altered as a result of the application of the technology.
5. Patient outcomes. Application of the technology benefits the patient.

The decision about how far down this hierarchy one needs to go before considering the clinical assessment complete will depend on a number of factors, including patient practice, the cost of technology, its toxicity, and one's values. The decision about how stringent to be in applying the preceding criteria is one of the major issues in the assessment of diagnostic technologies. Because this decision has a major influence upon the design of studies to assess diagnostic technologies, we will discuss these at length in Chapter 6.

Step 4: Community Effectiveness

Once a therapeutic technology has been developed (that is, once it leads to improved health outcomes) and is known to be efficacious, or if a diagnostic technology leads to better identification of those individuals who would benefit from efficacious therapy, information needs to be collected on how well this new technology will work in the field. The determination of community effectiveness involves the assessment of the magnitude of health improvement that can be expected because of the application of the specific technology in the community or population of interest. It may be defined as follows:

community effectiveness = efficacy x diagnostic accuracy x health professional compliance x patient compliance x coverage

We have discussed obtaining the information on the first two components; we will now address the remaining three.

Evaluation of Health Professional Compliance

This component of community effectiveness requires information on the extent to which health professionals comply with the requirements for application of appropriate diagnostic and management technologies (prevention, cure, palliation, and rehabilitation) (Department of Clinical Epidemiology and Biostatistics, McMaster, 1984a). In the literature on quality of care, actions of the health provider are often called "clinical process" (to differentiate them from "structure," which refers to the supply of facilities and qualified professionals, and from "outcomes'" which refers to the patients' health status).

That surgeons have different degrees of expertise has been blamed for the variation in peri-operative mortality rates of surgery technologies, such as cardiac valve replacement and coronary artery bypass. Evidence suggests that infrequent use of a technology by health care professionals is associated with poorer outcomes (Luft, Bunker, and Enthoven, 1979). Continuing medical education is important for ensuring that physicians and the other professionals involved correctly apply new technologies, for example, the new vasodilator drugs for cardiac failure and aspirin for transient ischaemic attacks. Some form of quality control may be necessary to ensure optimal application of new technologies with CT scans and radiotherapy devices.

Evaluation of Patient Compliance

The methods for assessing the extent to which patients comply with the health providers' recommendations and treatment are well described by Haynes and colleagues (1979). Home monitoring of blood glucose by pregnant diabetic patients and the care with which a sterile technique is followed for home renal dialysis are examples of situations in which poor patient compliance with implementation of diagnostic or therapeutic technologies can substantially influence the magnitude of benefit obtained from them.

Evaluation of Coverage

"Coverage" here is defined as the extent to which an efficacious technology is being appropriately applied to all patients or citizens who could benefit from it. Coverage should be differentiated from patient compliance; coverage describes whether or not the individual patient in need of the technology makes contact with the health professional who can provide the technology. Health professional compliance encompasses the subsequent clinical diagnosis and management actions by the health professional.

Information about coverage requires that the need for health care technologies be compared with their actual use in a defined population during a specific period. Appropriate utilization of health care technologies is influenced by (1) their accessibility, (2) awareness of their availability, (3) their acceptability, and (4) their linkage with the other components of efficacious care.

Accessibility is usually assessed by estimating the supply of services (the ratio of technology resources to population), taking into account the distance between the technology resource and those in need. Planners of ambulance services, for example, use studies showing the time that elapsed between receiving the request for an ambulance to delivery of the patient to a hospital. Technologies that require the patient's presence (as opposed to a blood sample), that are required for medical emergencies, that require repeated application, or that are directed primarily at diagnosis usually need to be more widely dispersed than elective major therapeutic technologies where patients can safely travel longer distances (for example, heart valve replacement and kidney transplantation).

Awareness of the availability of technologies by both health providers and patients is important. Health professionals are not always aware of the availability of new technologies. Photocoagulation therapy for diabetic retinopathy is of demonstrated benefit and is now widely available, but many physicians were unaware of this eighteen months after good controlled trials were published (Stross and Harlan, 1979).

Many technologies are painful or frightening to patients, that is, "unacceptable." Any procedure that is invasive can be painful, and injection of contrast media is still needed for many new technologies, such as CT scanners. The nuclear magnetic resonance scanners are noisy and can make the patients claustrophobic. CT scanners housed in cancer hospitals may frighten away patients.

Technologies for screening, such as blood pressure tests, cervical smears, and mammograms, need to be integrated into the health care system so that the patient is treated rather than neglected or just labelled. The identification of abnormalities suggestive of a urinary infection from screening of the urine of hospitalized patients is often not followed up (Williamson, Alexander, and Miller, 1967). As another example, hypertension screening programs administered in shopping plazas run the risk of informing the patients that they have hypertension without providing appropriate treatment (Silverberg et al., 1974).

Predicting the Magnitude of Community Effectiveness

All five components of community effectiveness can be multiplied together to estimate the predicted magnitude of effectiveness of a new medical technology in the community. Community effectiveness can be expressed as a percentage of the burden of illness dealt with, using the appropriate indices (increases in symptom-free days, decreases in mortality or disability days).

The best evidence on the combination of all of the components of community effectiveness (except for coverage) comes from randomized management trials where the new technology is compared with the existing technology in current use in a practical setting. Examples of such studies include the studies of hypertension in steel workers in Canada (Haynes et al., 1976) and that by the Hypertension Detection and Follow-up Cooperative Group (1979). Follow-up studies in the United States include that by the Veterans Administration Cooperative Study Group (1970), the coronary artery bypass studies (Cass, 1983a, 1983b), the studies of coronary care units in England (Hill, 1973), and the studies of the effectiveness of nurse practitioners in primary care in Canada (Spitzer et al., 1974). These studies are, unfortunately, uncommon. It is also likely that compliance by health providers and by patients may vary from region to region and from country to country so that caution should be exercised before generalizing the results. In contrast to management studies, the estimates of efficacy from the results of exploratory trials appear to be more stable, and it is probably unnecessary to repeat efficacy trials in most situations once two or more well-designed trials give consistent results.

Coverage is best assessed by surveys that sample the appropriate population for asymptomatic diseases. This may require a community survey, or appropriate alternatives. Where the condition is severe enough for patients universally to seek help from health professionals or to be admitted to hospital, then it is far cheaper to survey these health professionals or hospitals.

For example, for conditions such as hypertension in situations where almost everyone attends their family physician regularly, a survey can be conducted by the family physician (Ontario Council of Health, 1977). The potential flaws of utilization data, already noted in Step 1 and discussed further in Chapter 8 for assessing the burden of illness, apply here.

Where trials of benefit in practical settings have not been done, community effectiveness has to be estimated indirectly by using data from exploratory trials and then taking into account the other factors; individual estimates of each of the five components of community effectiveness should be made. For components for which the data is inadequate, appropriate studies should be carried out (Tugwell et al., 1985). This estimate of community effectiveness is needed for the analyses of cost effectiveness described in the next step of the loop.

Step 5: Efficiency

The preceding 3 steps of the loop have addressed whether a new health care technology does more good than harm when applied in a clinical setting. This step provides information on whether the technology is being delivered to those who would benefit from it with an optimal use of resources and involves the relationship between costs and effects.

$$\text{efficiency} = \text{patient benefit (outcome) / cost} \quad \text{or} \quad \text{net costs(\$), that is costs} - \text{benefits}$$

Efficiency is expressed as the effects (number of lives saved, number of disability days avoided) obtained for a specific *cost* (expressed in dollars). Again, the assessment of efficiency should not be conducted in the absence of evidence of efficacy and community effectiveness unless, of course, both are studied concurrently. Other issues that should be considered in the assessment of efficiency have been discussed by Stoddart and Drummond (Department of Clinical Epidemiology and Biostatistics, McMaster, 1984b). The following three approaches are those used most widely.

1. Cost-Effectiveness. Effectiveness in this approach is measured in a common unit of health, such as lives saved, levels of function, proportion of patients in whom symptoms are controlled. The measure may be derived from *efficacy* studies in a tightly controlled situation or from studies incorporating variable numbers of the other components that compose *community effectiveness*.

2. Cost-Benefit. Effectiveness (benefit) in this approach is measured in the appropriate unit of health and then converted into monetary units, that is, dollars.
3. Cost-Utility. Effectiveness (utility) in this approach is measured in the appropriate unit of health, such as lives saved, functional improvement, symptom control, and then converted into "utility units" that are measures of the relative social value (importance) of the outcomes.

Many of these issues will be explored further in Chapter 8 (on optimal facilities planning) but it would also be beneficial if economic evaluations incorporated consideration of them as *part of evaluation methodology*. The sensitivity analysis approach discussed in Chapter 7 has been used constructively by some analysts in this regard.

Identifying Value Judgements. It should be clear from the above discussion that economic evaluation methodology involves a number of value judgements, including the choice of discount rate and the methods of valuing the costs and benefits. It is important that such judgements be made explicit when results are presented, so that decision makers acting on the results of the study do not commit themselves to these unknowingly. Indeed, the strength of a good economic evaluation would be that value judgements are made more explicit than in alternative ways of arriving at a decision, such as "professional judgement."

As with some of the uncertain parameters, the analyst may often undertake a sensitivity analysis of the results to factors such as discount rate.

Recognition of Other Factors in Decision Making. The better economic evaluations will allude to other relevant factors in decision making in health care. A major factor is *equity* in the provision of services, for instance by geographical area or income class. In some situations the decision maker might make a conscious trade-off between equity and efficiency. For example, a policy for locating high technology diagnostic and therapeutic facilities based on equalizing patient travelling time may be at odds with one based on effective and efficient utilization of those facilities. The economic evaluation can, however, estimate the cost (in terms of extra resources that could be used elsewhere) of such an equal access policy.

Step 6: Synthesis and Implementation

This step integrates feasibility, estimated community effectiveness, and efficiency to make recommendations for diffusion of technologies or, for existing technologies, appropriate utilization. To make such recommendations, one must lay out the advantages and disadvantages of the different options. It is important to consider the coverage of existing facilities, the specifications for assessing the feasibility of different approaches to optimal facilities planning, and the quantitative estimates of community effectiveness and efficacy. Cultural, socio-economic, and political factors will affect the feasibility of different diffusion options and thus may influence policy discussions. The different diffusion models were described in Chapter 3. The prescriptive counterpart will be addressed in Chapter 8, on facilities planning. (Chapter 8 describes an established approach from business location theory that maximizes the total benefits to the community of a new or current technology and takes into account the current and future size, number, and location of institutions and health professionals.) Step 6 then requires information of adequate quality for burden of illness, health needs, community effectiveness, and economic efficiency derived from the studies described in the preceding steps of the loop.

Step 7: Monitoring and Reassessment

This step sets targets for benefits and costs, assesses the implementation of the diffusion of the technology, and assesses the success of the technology in reducing the burden of illness. Setting targets is important both for defining realistic objectives and for assessing success. Targets for benefits and costs will usually have been set during the decision making of Step 6 when the potential benefit for a specified cost of the different diffusion options for each technology were compared.

The assessment of the implementation of the diffusion of the technology needs to be tailored to the individual technology and may consist of short-term, intermediate, and long-term criteria of success that make up a profile of markers selected to indicate prognosis. These monitoring methods usually focus on one or more of the following categories (Department of Clinical Epidemiology and Biostatistics, McMaster, 1984a):

1. structural considerations, such as buildings equipped with the technology, for example, operating rooms appropriately equipped for coronary artery bypass, rooms with magnetic protection for an NMR
2. evidence of satisfactory training of personnel to operate or apply the new technology, for example, surgeons trained to perform coronary artery bypass, technicians trained to operate the NMR
3. evidence of appropriate utilization of the technology by demonstrating that the expected number of patients with the clinical problems benefited from the technology, for

example, evidence that an expected number of patients with unresponsive angina, with proven stenosis of the left main coronary artery greater than 75 per cent, underwent coronary artery bypass by a specified date combined with the evidence of a sample survey of cardiologists, and/or primary care physicians' charts in the area to ensure that appropriate patients are being offered the technology

4. indicators of patient outcomes, for example, review of charts of patients who die in the region to see whether the possibility of coronary artery bypass was offered to patients with unresponsive angina without contraindications.

Finally, assessment of the reduction of the burden of illness in the population of interest, using the same methods as for Step 1, is important to "close" the technology assessment cycle by returning again to the residual, unmet health needs. The cycle is then repeated to provide the information for continuing technology-funding decisions.

Conclusion

As is the case for much of this book, this chapter presents a somewhat idealized vision of health care planning. The loop is intended to provide a set of scientific guidelines that can be useful as a framework for identifying some of the important factors and for assessing the quality of information used in making decisions about health. It is fully appreciated that only rarely (or never) will ideal information be available. In the context of incomplete data, the loop and the associated guidelines, which upcoming chapters shall delineate, provide a framework for organizing available information systematically and identifying potential flaws explicitly. We believe that awareness of the quality of the evidence supporting the need for, and the effectiveness of, a new technology will allow health care providers and health policy makers to make wiser decisions on how health technologies should be disseminated and used.

Another caveat should, however, be introduced here, lest the authors appear hopelessly naive. If funds are withheld from dissemination or use of a technology for which there is minimal need or marginal effectiveness, will these funds be appropriated for the next best cause? For this to happen would require trustworthy health status measures, agreed-upon priorities, and, above all, rational centralized planning. Ongoing work has generated high-quality health status measures, and further improvements are on the horizon; the prospect of agreed-upon priorities is, however, more remote. Finally, in a country where production of tobbaco, the primary preventable cause of morbidity and mortality, is subsidized by the government, it would be understandable if the reader despaired of the third criterion being realized. Setting priorities is and should be part of the political process. The process of ensuring that these priorities are implemented rationally may be undercut by political contingencies. We nevertheless believe that providing the tools for rational decision making will improve the quality of the decisions made. The loop, and the book in general, are an attempt to provide a framework we should strive to realize, even if the realization will be, at best, imperfect.

References

Canada, Health and Welfare Canada. June 1981. *The Health of Canadians: Report of the Canada Health Survey*, Catalogue 82-538E. Ottawa: Statistics Canada.

Cass, Principal Investigators and Their Associates. 1983a. "Coronary Artery Surgery Study (CASS): A Randomized Trial of Coronary Artery Bypass Surgery." *Circulation* 68: 939-950.

Cass, Principal Investigators and Their Associates. 1983b. "Coronary Artery Surgery Study (CASS): A Randomized Trial of Coronary Artery Bypass Surgery – Quality of Life in Patients Randomly Assigned to Treatment Groups." *Circulation* 68: 951-960.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1984a. "About the Quality of Clinical Care." *Canadian Medical Association Journal* 190: 332-381.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1984b. "How to Read Clinical Journals VII: To Understand an Economic Evaluation." *Canadian Medical Association Journal* 130: 1428-1434, 1542-1549.

Haynes, R.B.; Sackett, D.L.; Gibson, E.S.; Taylor, D.W.; Hackett, B.C.; Roberts, R.S.; Johnson, A.L. 1976. "Improvement of Medication Compliance in Uncontrolled Hypertension." *Lancet* 1: 1265-1268.

Haynes, R.B.; Sackett, D.L.; Taylor, D.W.; Gibson, E.S.; Johnson, A.L. 1978. "Absenteeism from Work Following the Detection and Labelling of Hypertension." *New England Journal of Medicine* 299: 741-744.

Haynes, R.B.; Taylor, D.W.; Sackett, D.L., eds. 1979. *Compliance in Health Care*. Baltimore: Johns Hopkins University Press.

Hill, J.D. 1973. "A Randomized Trial of Home Versus Hospital Management for Patients with Suspected Myocardial Infarction." *Lancet* 1: ;837-841.

Hypertension Detection and Follow-up Program Cooperative Group. 1979. "Five-year Findings. I: Reduction in Mortality of Persons with High Blood Pressure, Including Mild Hypertension." *Journal of the American Medical Association* 242: 2562-2571.

Kohn, R. and White, K.L., eds. 1978. *Health Care: An International Study*. Toronto: Oxford University Press.

Luft, H.S.; Bunker, J.P.; Enthoven, A.C. 1979. "Should Operations Be Regionalised. The Empirical Relation between Surgical Volume and Mortality." *New England Journal of Medicine* 301: 1364-1369.

Ontario Council of Health. 1977. *Hypertension*. Report of the Task Force on Hypertension. Toronto Ontario Council of Health.

Roos, N.P., and Roos, L. 1982. "Surgical Rate Variations." *Medical Care* 20: 945-958.

Silverberg, D.S.; Smith, E.S.O.; Juchli, B.; Vandorsser, E. 1974. "Use of Shopping Centres in Screening for Hypertension." *Canadian Medical Association Journal* 111: 769-774.

Spitzer, W.O.; Sackett, D.L.; Sibley, J.C.; Roberts, R.S.; Gent, M.; Kergin, D.J.; Hackett, B.S.; Olynich, A. 1974. "The Burlington Randomized Trial of the Nurse Practitioner." *New England Journal of Medicine* 290: 251-256.

Stason, W.B., and Weinstein, M.B. 1977. "Allocation of Resources to Manage Hypertension." *New England Journal of Medicine* 298: 732-739.

Stross, J.K., and Harlan, W.R. 1979. "The Dissemination of New Medical Information." *Journal of the American Medical Association* 241: 2622-2624.

Tugwell, P.; Bennett, K.J.; Sackett, D.L.; Haynes, R.B. 1985. "The Measurement Iterative Loop." *Journal of Chronic Disease* 38: 339-351.

Veterans Administration Cooperative Study Group on Antihypertensive Agents. 1970. "Effects of Treatment on Morbidity in Hypertension." *Journal of the American Medical Association* 213: 1143-1152.

Williamson, J.M.; Alexander, M.; Miller, G.E. 1967. "Continuing Education and Patient Care Research." *Journal of the American Medical Association* 201: 118-122.

Chapter 5:

Guidelines for Health Technology Assessment: Therapeutic Technologies

Gordon Guyatt, Michael Drummond, David Feeny,
R. Brian Haynes & Peter Tugwell

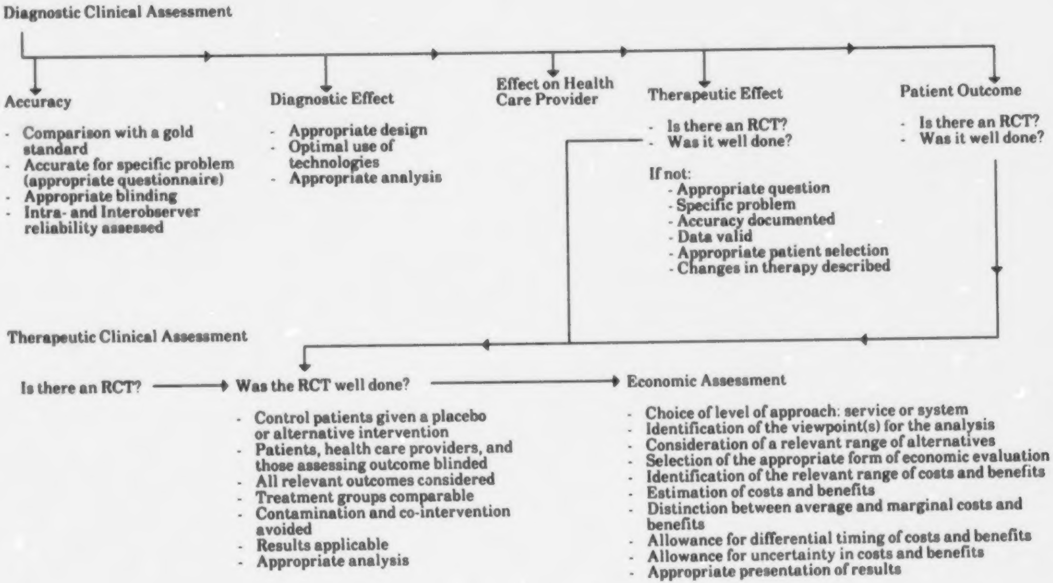
The strategies for definitively establishing the efficacy of a technology (for a specific condition in a specific group of patients) have become well established through repeated and convincing use in trials of pharmacologic agents. Although it has been strongly argued that the same principles are applicable to other therapeutic technologies, such as surgical interventions, (Spodick, 1982) this viewpoint is not universally accepted (Bonchek, 1979). We believe that the guidelines that follow are applicable to any therapeutic technology, including drugs (both those developed and administered in the traditional way as well as innovative techniques, such as the portable insulin pump) as well as transluminal angioplasty, implantable cardiac defibrillators and intestinal pacemakers, radiotherapy, bone marrow transplantation, plasmapheresis, new operations, educational programs, and multifaceted interventions such as rehabilitation, microsurgery, cerebral bypass surgery, knee replacement programs, and intensive care units. We have summarized our criteria in Figure 5.1 and describe them more fully in the following sections.

Was the Appropriate Question Framed?

One cannot simply talk about whether a particular treatment is any good; one must specify good for doing what, to whom, and compared with what. For example, coronary artery bypass grafting (CABG) is increasingly advocated for patients with coronary artery disease without appropriate definition of their clinical characteristics. Patients operated on include those with stenosis of the mainstem of the left coronary artery, those with angina, those with acute myocardial infarction, and asymptomatic patients with high levels of risk. Yet, the only good evidence shows that it increases longevity only for the first of these groups (Cass, 1983a). The potential for doing more harm than good recommending CABG to asymptomatic patients, as well as the immense resources required if CABG were offered to

FIGURE 5.1

Guidelines for Assessment of Technology



Diagnostic Clinical Assessment

Accuracy

- Comparison with a gold standard
- Accurate for specific problem (appropriate questionnaire)
- Appropriate blinding
- Intra- and Interobserver reliability assessed

Diagnostic Effect

- Appropriate design
- Optimal use of technologies
- Appropriate analysis

Effect on Health Care Provider

Therapeutic Effect

- Is there an RCT?
- Was it well done?

If not:
- Appropriate question
- Specific problem
- Accuracy documented
- Data valid
- Appropriate patient selection
- Changes in therapy described

Patient Outcome

- Is there an RCT?
- Was it well done?

Therapeutic Clinical Assessment

Is there an RCT? → Was the RCT well done? → Economic Assessment

Was the RCT well done?

- Control patients given a placebo or alternative intervention
- Patients, health care providers, and those assessing outcome blinded
- All relevant outcomes considered
- Treatment groups comparable
- Contamination and co-intervention avoided
- Results applicable
- Appropriate analysis

Economic Assessment

- Choice of level of approach: service or system
- Identification of the viewpoint(s) for the analysis
- Consideration of a relevant range of alternatives
- Selection of the appropriate form of economic evaluation
- Identification of the relevant range of costs and benefits
- Estimation of costs and benefits
- Distinction between average and marginal costs and benefits
- Allowance for differential timing of costs and benefits
- Allowance for uncertainty in costs and benefits
- Appropriate presentation of results

all patients with angina, emphasizes the importance of framing the appropriate question.

Was the Design of the Studies Sufficiently Rigorous for Making Policy Decisions?

Controlled trials of all new therapeutic technologies that are potentially harmful or whose diffusion would involve major costs should, in most cases, be carried out to assess benefit before decisions regarding diffusion are made. Information on toxicity from these controlled trials frequently needs to be supplemented by information from additional studies to detect rare but dangerous side effects. We will consider first assessment of benefit, then methods to evaluate toxicity.

Study Designs for Assessing Benefit

Before controlled trials (the definitive experiments conducted to determine a technology's effectiveness) are begun, preliminary studies are generally undertaken to delineate the specific conditions, and classes of patients, for which the treatment is likely to provide benefit. These studies usually have before-after designs. In the new technology of coronary artery transluminal angioplasty, for example, a consecutive series of patients with intractable angina undergo coronary artery angioplasty and are followed to see how many improve. The controls in these studies, if any, are historical; that is, the results with coronary artery angioplasty are compared with how patients with intractable angina have fared in the past. Except under extraordinary circumstances (such as a striking reduction in the mortality of what was formerly a uniformly fatal disease), these types of studies cannot provide definitive answers about whether a technology is effective, but they can tell us the conditions and the patients in which the treatment is promising enough to make further studies worthwhile.

In addition to providing preliminary clues about ultimate effectiveness, research in this phase can delineate potential toxicity that would make further studies pointless or unethical. There are many examples of drugs and machines for which the initial enthusiasm faded quickly as a result of troubling or dangerous side effects. Clearly, providing a preliminary indication of a technology's toxicity is a major goal of studies conducted to define potential applications.

Chalmers (1975, 1982) has made cogent arguments to "randomize the first patient." The suggestion that we dispense with the phase of defining potential applications for therapeutic technologies is, at least in part, a response to instances in which large-scale studies have been done, still leaving doubt about whether a therapy is worthwhile. Another danger of persisting too long with exploratory studies is that physicians may be reluctant to randomize their patients if they have a strong (or even a not-so-strong) suspicion that a treatment will favourably influence outcome. This problem is most germane for the cancer therapies. In an otherwise untreatable and uniformly fatal condition, physicians may consider it unethical to withhold a potentially effective therapy. However, the controlled trials necessary to provide compelling evidence of effectiveness are expensive, logistically difficult, and generally very time consuming. In most situations, a technology can be rejected for a specific indication in a specific group of patients if preliminary research shows no hint of possible improvement in outcome, or if prohibitive toxicity turns up, thus sparing the effort and expense of definitive studies.

The answer to these problems is not to abandon the exploratory stage, but rather to limit its scope and duration. The timing of the first controlled trial will depend on the technology under consideration, the results of animal work, the likelihood of substantial toxicity, and the rate of spontaneous remission in diseased patients. Occasionally, the benefit of a new technology in the stage of defining potential applications will be so dramatic that no further studies are necessary.

We will examine in further detail in subsequent sections, situations where controlled trials are unnecessary, but first we will review the major strengths of controlled trials for assessing the benefit of new health care technologies.

Justification of Insistence upon Controlled Trials for Assessing Benefit for Most Technologies

Randomized controlled trials (RCTs) are expensive, logistically difficult to conduct, and, for technologies other than drug therapy, their preeminence as a method for assessing effectiveness is still controversial. Will we really go so terribly wrong if we rely on studies of inferior design? There are many examples of therapies that have received enthusiastic ratings on the basis of non-RCT evidence but that have proved to be useless, or even harmful, when definitive studies were undertaken. Two particularly infamous examples are gastric freezing for gastrointestinal bleeding (a procedure in which a balloon attached to a tube was inserted in the stomach and irrigated with alcohol cooled to subfreezing temperatures), and mammary artery ligation for coronary artery disease. On the basis of uncontrolled studies reporting striking success both these technologies were fairly widely disseminated (Adams, 1958; Ellis et al., 1958). Eventually, however, RCTs were done that proved the technologies worthless (Cobb et al., 1959).

Two other less infamous, but equally instructive, examples are the use of hydralazine as a treatment for chronic heart failure and the use of intermittent, positive pressure breathing as prophylaxis against post-operative respiratory complications. Hydralazine is a vasodilator effective in the treatment of hypertension, which, a few years ago, was suggested as a potential therapy to relieve symptoms in patients with chronic heart failure. The drug was shown to improve acute hemodynamic measurements in heart failure patients, and in uncontrolled studies appeared to ameliorate symptoms (Chatterjee et al., 1976; Franciosa, Pierpont, and Cohn, 1977; Rubin et al., 1979). On the basis of this information, hydralazine became popular in the treatment of heart failure. However, a recent randomized control trial showed no benefit whatsoever (Franciosa et al., 1982). A device developed to deliver intermittent positive pressure to prevent post-operative respiratory complications was widely disseminated but suffered a similar fate when a number of controlled trials showed that it was ineffective; it has since been largely abandoned (Baxter and Levine, 1969; Dohi and Gold, 1978).

Further evidence that inferior designs should not be relied upon for assessing benefit is contained in a literature review by Sacks, Chalmers, and Smith (1983). These authors examined fifty controlled trials of six different therapies and compared them with fifty-six studies of the same treatments in which inferior designs were used. Whereas 79 per cent of the studies using inferior design found the therapy better than the control regimen, only 20 per cent of the controlled trials agreed.

Fortunately, these experiences have led to more cautious evaluation of some therapeutic technologies. For example, extremely enthusiastic reports suggesting that hemodialysis was of benefit to almost all patients with schizophrenia led not to widespread adoption of the technology for this condition, but rather to a controlled trial that demonstrated no effect.

We will now review some of the issues involved in the conduct of randomized controlled trials of therapeutic technologies.

Importance of a Control Group in Study Design

A simple before-after study takes a group of patients who are ill, applies an intervention, and monitors them to see if they improve. Unfortunately, improvement in such a study cannot be necessarily attributed to the intervention. Other factors such as the placebo effect associated with institution of treatment, the attention the patients received, other concomitant change in therapy, bias in measurement, and the natural history of the illness (that is, they would have improved anyway) can all explain the improvement in a before-after study.

The solution to this problem is to include a control group – patients who are, within the limits of feasibility, identical to the experimental group but who do not receive the therapeutic technology – and to follow the two groups over time. At the end of the investigation one then compares those who received the technology with those who did not (either in absolute terms or according to how much they changed during the course of the study). This study design allows considerably more confidence that differences between the two groups are indeed due to the intervention.

The absence of a control group leads, in general, to overly sanguine assessments of a technology's potential benefit. This is not really very surprising: the biases (by which we mean the factors that lead to findings systematically different from the truth) inherent in uncontrolled studies all tend to favour the experimental treatment. Investigators tend to choose subjects to receive an as yet untested treatment at the low ebb of their illness. This means that the patients chosen are likely to get better on their own, even without the new technology. Any placebo effect (the improvement that is almost universally seen after institution of a new treatment, even if it has no biological effect on the underlying disease) will be attributed to the technology. Both the patients themselves and the investigators would like to see the new treatment work and will therefore tend to err on the side of improvement when they evaluate whether any change with therapy has occurred. Given these factors, it is not surprising to see findings such as those of Sinclair, who evaluated the literature regarding the treatment of respiratory distress syndrome in infants and found that the reported success in uncontrolled studies was 85 per cent, but the comparable figure for controlled trials was only 50 per cent (Sinclair, 1966).

Selection of a Control Group

Having decided that a control group is necessary to ensure the valid assessment of a new technology, one has to decide on how the control group should be selected. The easiest way is to review records of patients treated before the advent of a new technology and to compare them with current patients receiving the experimental treatment. Unfortunately, there is no way of guaranteeing the comparability of the treated and control patients when this sort of comparison is used. Any difference between treated subjects and historical controls may be due to changes in the patient population (current patients may not be as sick), changes in the natural history of the underlying condition, changes in management independent of the technology under investigation, or differences in the method of measuring outcome. That such factors are extremely important was demonstrated in the literature review by Sacks, Chalmers, and Smith (1983), referred to earlier, where 79 per cent of the studies using historical controls found the therapy better then the control regimen, and only 20 per cent of the randomized studies agreed. One strategy that has been suggested to deal with the lack of comparability of historical controls is to statistically adjust the results for prognostic factors known to affect outcome. It is argued that once this is done, the comparability of the historical controls and the experimental group can be assumed. We generally know about, and can measure, however, only a small proportion of the factors that determine outcome. Given this limitation, it is not surprising that Sacks and his colleagues found that adjusting the outcomes of the historical controls for prognostic factors did not appreciably change their results.

Given that we must use concurrent control groups to ensure the validity of studies investigating the utility of therapeutic technologies, how should patients be chosen to receive the new technology or the control treatment? By far the best method is randomization, a strategy comparable to tossing a coin, in which each subject has an equal chance of being allocated to the experimental or the control group. The major advantage of randomization is that not only are known prognostic factors likely to be distributed equally between the two groups, but those factors that influence outcome but that we do not know about, or cannot measure, will have a similarly even distribution. In addition, conscious or unconscious bias, which can come into play when an investigator decides, by whatever criterion, which patients enter the treatment and control groups (so that, for example, healthier patients are allocated to receive the experimental technology) is eliminated. Therefore, trials in which patients are allocated according to a randomization schedule – randomized control trials – provide the most valid assessment of a technology's effectiveness.

Placebo Versus No Intervention for Control Group

A placebo can be defined as any treatment or aspect of a treatment that does not have a specific action on the patient's symptoms or disease (Benson and McCalla, 1958). The placebo effect is involved whenever patients are offered therapy of any sort. Even if the therapy does have a specific action on the disease, non-specific factors will influence the patient's response. These include the form of the intervention (whether a pill, a surgical procedure, or a stroke rehabilitation program), the interaction with the health care provider, and the therapeutic setting. The placebo effect is extremely powerful: for most interventions one can expect that about one-third of patients will improve on the basis of the placebo effect alone (Beecher, 1955). When an enthusiastic and trusted physician presents the interventions, the placebo response rate may increase to around 70 per cent. Perhaps the most dramatic demonstration of the placebo effect in medical history occurred in two controlled trials of internal mammary artery ligation (Cobb et al., 1979; Diamond, Kittle, and Crockett, 1960). In these studies, the investigators subjected their placebo group to a sham operation and found that symptom relief occurred in more than 75 per cent of both active treatment and control groups; and comparable reductions in nitroglycerine use, increased exercise capacity, and improvement in exercise electrocardiograms were also found. Given the remarkable power of the placebo effect, it is no wonder that so many uncontrolled studies provide gross overestimates of treatment effectiveness.

What would have happened if the scientists investigating internal mammary artery ligation had conducted a randomized control trial, but instead of including the sham operation had randomized their patients to surgery or no new therapy whatsoever? They would have found that more than 75 per cent of the patients who were operated on, and a negligible proportion of the controls, improved over time. They would have then concluded that internal mammary artery ligation is a remarkably effective treatment for angina; and, perhaps contrary to one's intuitive impression, they would have been right! That is, taking the patient to the operating room, ligating the internal mammary artery, and when meeting them thereafter maintaining an optimistic attitude about the likelihood of improvement probably does make patients feel better, improve their ability to exercise, decrease their use of nitroglycerine, and even improve their exercise electrocardiograms. The effect is not, however, specific to the disease, but due to the power of the intervention as a placebo.

The underlying issue of whether an effective treatment is working for the reasons the investigators think it is working has been called "construct validity." The previous example demonstrates the importance of knowing why a treatment is effective; an otherwise perfectly conducted RCT of internal artery ligation without the placebo operation could have led to the widespread dissemination of what is (if one excludes its power as a placebo) an ineffective treatment. Pills, while probably not as potent placebos as operations, are imbued with substantial power to yield nonspecific improvement. For this reason it is virtually universally accepted that to ensure the validity of an RCT testing a new drug, one must administer an identical placebo to the control group.

The issue with non-drug technologies is more difficult, for it is often impossible to find an adequate placebo treatment. Sham operations, such as those used to investigate internal mammary artery ligation in the late 1950s, would not be ethically acceptable in the 1980s, and this situation is unlikely to change. The same ethical constraints apply to other procedures such as mock dialysis or mock transfusions of blood products. The result of this problem is situations such as the one we are facing with coronary artery bypass grafting for patients with coronary artery disease: though in many patients the surgery does not prolong life, it is more effective than medical therapy in decreasing symptoms (Cass, 1983b). Whether this symptomatic benefit is due to improved cardiac blood supply attributable to the surgery, or to the fact that surgery is a more potent placebo than is medical therapy, we do not know.

Because there are better ways of providing effective placebos than expensive and risky surgery, the absence of adequate placebos for many therapeutic technologies is a major problem. At least part of the solution lies in the choice of outcome measures. While one might expect the placebo effect of CABG to improve symptoms, to decrease nitroglycerine use, and perhaps even to improve objectively measured exercise capacity, one would not expect it to prolong life. This implies that mortality is an outcome that is probably free from placebo influence. The message is that when adequate placebos have not been used, one can only trust "hard" outcomes, such as mortality, on which it is unlikely that the placebo effect of the technology will have a major effect. Although other, "softer" outcomes such as symptoms should be measured, attributing any differences between treatment and control groups to a disease-specific effect of the technology should only be done after a careful consideration of the possibility that the differences are due to a placebo effect.

Surgical technologies are not the only instance in which it is difficult to ascertain which aspect of treatment has changed outcome. Difficulties arise, for example, in multifaceted technologies such as interdisciplinary rehabilitation programs for patients with chronic chest disease (Royal College, 1981). If patients are randomized to receive an in-patient rehabilitation program or continue with routine care, how can we know whether any difference in outcome is due to the hospitalization alone, the attention the patients received, interaction with other patients, or the educational, physical training, or psychosocial aspects of the program? The answer is, we cannot. There are three possible solutions to this problem. One is to evaluate components of the program separately: randomize patients to the whole program or the whole program without, say, the physiotherapy component. Alternatively, one might provide placebo treatment, such as hospitalization, without any other aspect of the program. Finally, one could use both alternatives together: patients could be randomized to receive the basic program with physiotherapy, or the basic program with an attention placebo. In any case, it is clear that with complex therapeutic technologies, care must be taken to determine what aspects of the intervention are producing any changes observed.

When Are RCTs Inappropriate for Assessing Benefit?

We have argued strongly that, before dissemination of therapeutic technologies, evidence of effectiveness obtained from RCTs is necessary. RCTs do not, however, constitute a panacea for the problems of technology assessment. For one thing, they are expensive and difficult to do well. Large sample sizes are often necessary to increase power to acceptable levels, and this may entail a multicentre study with its attendant logistic and administrative difficulties. Although these problems can be surmounted, there are other, inevitable limitations to RCTs. RCTs can be done only on patients who provide informed consent, and there is little doubt that such individuals differ systematically from the rest of the patient population. Therefore, extrapolating results from RCTs to all patients is, to some extent, a leap of faith. Negative RCTs in heterogeneous populations cannot exclude the possibility that although the group as a whole did not benefit, there is some small subgroup for whom the technology is effective. Finally, sample sizes sufficient to mount RCTs with acceptable power are not feasible in very rare diseases.

The expense and difficulty of doing RCTs, and their inherent limitations, argues for selective application. Are there any guidelines we can suggest for deciding when RCTs of therapeutic technology can be dispensed with? It has been suggested that if an amputee is given a potion and five minutes later has regrown a new limb, an RCT of the novel treatment is not required. Although we do not have any therapies that are quite that dramatic, there are examples of situations in which definitive demonstration of effectiveness did not require an RCT. The common denominator of many of these situations is that before the development of the new technology the condition was uniformly, or almost uniformly, fatal. This is true for dialysis in end stage renal failure, cyclophosphamide in Wegener's granulomatosis, and insulin in diabetes. Other therapies work quickly and dramatically enough that case series or before-after studies are all that are required for demonstration of effectiveness: examples include penicillin for pneumococcal pneumonia, chemotherapy for lowering the number of abnormal cells in leukemia, and joint prostheses for destroyed hip joints.

It may be appropriate to forgo an RCT when a technology is very low in cost and devoid of toxicity. In this situation, the cost of the studies needed to prove effectiveness may be simply not worth it. Finally, one might hesitate to conduct an RCT of a therapeutic technology if the results will not affect its utilization.

Toxicity

The demonstration of effectiveness does not by itself indicate clinical applicability of therapeutic technology. Technologies may be effective, but accompanied by unacceptable side effects. Defining the limits of a technology's toxicity requires, by and large, different strategies from those required for documentation of effectiveness.

RCTs are appropriate for discovering the frequency of common side effects. In fact, they are the only way of accurately defining the extent to which the most frequent side effects are specific to the drug. The reason is that there is a strong tendency in people to attribute the everyday ailments that afflict us all (headaches, stomach aches, gastrointestinal upset, and so on) to a new medication. In addition, just as there is a powerful positive placebo effect, placebos may be associated with their own toxicity. Therefore, the only way of definitively determining technology-specific toxicity is to compare the incidence of side effects in the treatment group with those experienced by patients taking the placebo; the difference between the two figures represents the morbidity that can be attributed to the technology itself.

There are, however, many ways in which RCTs are limited in their ability to delineate toxicity. As mentioned previously, patients who enter RCTs differ systematically from the population from which they are drawn. As a rule, they tend to be healthier, are not afflicted to as great an extent with diseases other than those under investigation, and may be on fewer other drugs. Each of these factors may reduce the toxicity of the technology in comparison with what would be found in a general population with the underlying disease of interest: healthier, low-risk patients may have a lower post-operative mortality associated with surgical technologies; other diseases may increase therapy-associated complications (increased infections in diabetics on dialysis); and drug interactions may enhance toxicity (for example, a toxic level of anticoagulants as a result of decreased metabolism secondary to concurrently administered medication).

Other problems associated with RCTs include limitations of size and duration. Rare side effects will not be picked up by RCTs designed to determine effectiveness, not toxicity. The number of patients required to be 95 per cent certain of seeing a single case of a particular toxic effect is three times the reciprocal of the incidence of the effect (Gent, 1984). For example, if a side effect occurs in one in a hundred people, 300 people are required to be 95 per cent certain that a single case of that side effect will be seen in a given study. This fact makes elucidation of toxicity beyond the scope of most RCTs intended to assess effectiveness when the frequency of toxicity is substantially less than that of the rate expected for positive benefits of the treatment.

A parallel problem exists with deleterious consequences associated with long-term use of therapeutic technology. Generally, effectiveness can be determined in weeks to months; toxicity that requires years to manifest itself will not be evident in short-term RCTs.

Other strategies are, therefore, needed for detailing a technology's side effects. One method is simply to report any odd events happening in patients taking a particular drug or exposed to a particular procedure. However, confidence in the causal connection between technology and a particular adverse effect requires that a number of conditions be met. These include (1) a strong and consistent association between the incriminated technology and the adverse effect, (2) an appropriate temporal relation (that is the technology is applied before the side effect appears), (3) a dose-response gradient (that is, the more of the technology one gets, the more likely is the occurrence of the side effect), and (4) the establishment of the biological link between the technology and the toxicity. Therefore, unless the untoward events are either frequent or almost never occur spontaneously or following other therapies (such as the characteristic limb deformities associated with thalidomide), it is difficult to know whether reported problems are really due to the technology. This is especially true when problems occur in sick patients with other conditions or sick patients treated with other drugs, in whom multiple plausible causes can explain the putative toxicity. These problems can be ameliorated to some extent by the use of registries, in which the problems and outcome of each patient treated with a new technology are reported to a central authority. Some investigators have suggested that registries can provide sufficient information to decide about effectiveness, though they share the limitations of any before-after study. They do, however, provide a mechanism of estimating the incidence of serious or highly technology-specific side effects. Post-marketing surveillance of new drugs provides another mechanism whereby unusual side effects can be identified and their frequency quantified.

More sophisticated study designs enable stronger inferences about the causal role of a therapeutic technology in producing specific side effects. Prospective cohort studies in which similar patients receiving and not receiving a technology are followed over time are almost as difficult to do well as are RCTs, and are rarely undertaken. An efficient strategy to provide some evidence about whether a particular untoward event is really causally related to a technology is to do a case-control study. In such a study one selects a number of patients with a particular problem (such as endometrial carcinoma) and an equal, or greater, number of matched controls. One then determines the proportion of cases and controls exposed to the putative causal factor (say, post-menopausal estrogens) and determines whether this factor was more frequent in the patients with carcinoma than in the controls. Case-control studies must contend with numerous potential sources of bias, and the selection of appropriate controls is both problematic and controversial. However, if a number of such studies find a high and consistent risk of a particular toxicity associated with a technology, it provides strong evidence of a causal relationship.

In determining efficacy other factors besides the basic study design can lead to false conclusions. We shall continue with a description of other criteria that can be applied to ensure that the results are both valid and applicable to health care.

Is Bias Avoided in the Intervention?

Bias can be introduced into the intervention as a result of contamination, co-intervention, or differences in compliance rates, any one of which can destroy the validity of the trial. Contamination occurs when the control group ends up receiving the active treatment. This may happen in drug trials, if patients allocated to placebo have access to the active drug; it has occurred in trials of CABG, in which patients allocated to medical treatment have opted, some time after randomization, for surgery. It is important to minimize this contamination and to analyse the result correctly if it occurs.

Bias due to co-intervention occurs when aspects of treatment other than the experimental intervention differ in the treatment and control groups. Co-intervention is common in unblinded studies in which the therapeutic technology prompts the physician to provide closer and more careful follow-up to patients receiving the experimental treatment. If such follow-up leads to changes or adjustment of other medications, or is in itself effective therapy (as a result of, for example, its placebo effect), or leads to more expeditious treatment of deterioration, a bias favouring the treatment group is introduced. In placebo-controlled studies, blindness to allocation of all those involved in the trial (patient, health professionals, and those administering measures of outcome) is important for minimizing bias from co-intervention.

Non-compliance, the failure of patients to take the treatment or of health providers to deliver the intervention in the prescribed manner, needs to be measured and minimized. Differing compliance rates between the experimental and the control groups will lead to bias, and inadequate compliance with an intervention that works will lead to a false negative result in an exploratory study.

Accurate Measurement of All Relevant Outcomes

Comprehensiveness

Ultimately, therapeutic technologies are designed to prolong the duration and improve the quality of people's lives. Wherever relevant, the effect of a technology on mortality, morbidity, and quality of life must be measured. We have pointed out that (at least if one is to be confident of the contribution of disease-specific mechanisms to therapeutic effect) technologies which must be studied without adequate placebos have to focus on outcomes which are unlikely to be influenced by the placebo effect. It is probably a much more common mistake, however, for modern medical scientists to focus on what is easily measurable at the expense of what is really important. The results of diagnostic tests have all too frequently been used as a substitute for truly relevent end points. For example, the research on vasodilators used for the treatment of heart failure has been characterized by treatment recommendations based on the effects of the technology on indices of cardiac function. Unfortunately, these measures of cardiac performance have proved to bear little relation to exercise capacity or functional status, and treatments initially touted as breakthroughs in the therapy of heart failure have proved useless, or of equivocal benefit, in long-term studies. These studies also emphasize that part of measuring outcomes appropriately is to ensure adequate follow-up. Some therapeutic technologies (such as educational programs, and stroke, respiratory, and cardiac rehabilitation programs) may produce apparent short-term benefit, only to show no difference in long-term outcome. Follow-up must continue, therefore, for as long as one wishes to be confident that the technology maintains its effect.

Another area in which all relevant outcomes have not been fully considered is that of cancer chemotherapy and radiotherapy, in which relatively little attention has been paid to measurement of quality of life. To date, the focus in cancer trials has rested on mortality and on disease-free survival. Given the profound reduction in quality of life that often attends the combined effects of therapy and the underlying disease, the lack of attention to measurement of quality of life may lead to dissemination and uses of cancer therapy that are not in the patient's best interest. This point is emphasized in the area of surgery for lung cancer in an article by McNeill, Weichselbaum, and Pauker (1978). The authors found that health professionals usually select the treatment that produces the longest five-year survival (in this case, lung resection), but a substantial proportion of patients selected radiotherapy, thus forgoing extra years of life several years down the road to avoid the possibility of death on the operating table. These considerations emphasize the need to measure quality of life and patient values and desires.

Accuracy and Responsiveness to Change

The methods selected for measuring the outcome of interest in benefit studies need to satisfy the following criteria if the results are to be meaningful: credibility, or face validity, and responsiveness to change.

First, the method chosen to assess the outcome of interest should *appear* sensible to the investigator and to the clinicians and policy makers who will apply the results. Outcome indices or scores requiring computation, complex graphic representations, or statistical transformations (McGill Pain Score; Minnesota Multiphasic Personality Inventory Scores) need to be interpretable by the consumer if the results are to be believed.

Second, the method or instrument chosen to assess the outcome of interest needs to be able to detect the smallest differences judged to be sufficiently important to influence the behaviour of those who will use the results. One of the reasons for the failure to include assessment of pain and quality of life in many studies, such as the original coronary artery bypass studies, was the absence of methods that would accurately detect improvement in these end points. Attention to minimizing the two main types of error, random and systematic error (or bias), is necessary to optimize the accuracy of any end point measure. The three major sources of variability that need to be minimized are: (1) that due to the patients (for example, inconsistent response of patients who are distracted, are tired, or have organic memory loss); (2) that due to the observer (for example, differences in assessment technique or inferring different conclusions from the same information on clinical outcomes whether obtained directly from the patient or from "hard-copy" such as X-rays or radionuclide scans); and (3) that due to instrument variability (for example, some instruments require careful calibration; different models of CT scanners have different abilities to detect small changes in the size of masses). Repeat observations in patients whose condition is known to be stable, the demonstration that the method can detect known changes in patients (such as in patients receiving an established efficacious intervention), and, wherever possible, assessments made without knowledge of the allocation (blind assessments) are needed to establish that the variability and bias of the method is minimized and that it is responsive to changes. Accurate and responsive outcome measures are essential to be sure that a negative result of an efficacy trial is due to the intervention being truly inefficacious and is not simply a result of the deficiency of the outcome measure.

Applicability

Experimental results are of maximum use if they can be easily applied in the clinical context. To ensure that this is the case, criteria for including subjects in (and excluding them from) a study must be fully documented. This includes a description of the referral base and an account of how, among all those eligible, the study subjects were chosen.

Another feature needed to ensure applicability is a description (in sufficient detail that it could be reproduced by others) of the technology being tested. This is particularly important for multifaceted interventions such as rehabilitation programs in which terms like physiotherapy or occupational therapy may mean different things to different people.

Appropriate Data Handling

Although some aspects of data handling appear inpenetrable to the non-statistician, the following five aspects are accessible to all.

1. Although randomization markedly increases the likelihood that prognostic factors will be equally distributed between treatment and control groups, it does not guarantee it. If determinants of outcome are unequally distributed between the two groups (that is, more of the sicker patients end up being given either active treatment or placebo), false positive or false negative results can ensue. There are a number of ways to handle this problem. If the trial is cross-over in design (that is, patients receive first one treatment option and then the other), balance between groups is not an issue (because the groups are identical). If one is not using a cross-over design, patients can be stratified according to factors known to influence outcome, and randomization from within strata will ensure balance for selected prognostic factors. Finally, one can examine the two groups according to determinants of outcome after the trial is completed and make a statistical adjustment for any imbalances that might exist. There are excellent statistical techniques for doing this, but clinicians are often not prepared to accept the results of a study where there is an imbalance of a major prognostic variable (Committee, 1975). It is advisable, therefore, to stratify for known prognostic variables whenever possible. Studies are weakened if the issue of possible imbalance is not considered and if the distribution of prognostic factors is not presented in the report of the experiment.
2. It is important to check that all randomized patients are included in the final analysis and are analyzed in the group to which they were allocated. When this is not done, major bias can result. For example, a controlled trial of carotid artery surgery versus medical therapy for transient ischaemic attacks (disturbances of the function of the brain due to temporary interruption of blood supply, often followed by complete strokes) found markedly improved outcome in surgically treated patients (Fields et al., 1970). Patients who suffered stroke or death at surgery or in the immediate post-operative period were, however, excluded. When these patients were included, it turned out that the risk of stroke or death was comparable in the two groups (Fields et al., 1970).

3. Errors in the strategy for analyzing data may be somewhat mysterious for the statistical layperson, but these, too, can unfortunately lead to invalid interpretation of study results. For example, a recent study of isosorbide dinitrate, a vasodilator designed to improve cardiac function in patients with heart failure, reported a statistically significant improvement in heart function in the treated but not in the control patients. However, because there was some improvement in the control group, when one compared the degree to which the two groups had improved over time (the appropriate comparison, and the very reason we bother with control groups), there was no statistically significant difference (meaning that the difference in improvement that was observed could easily have happened by chance). In this situation, an apparently positive result turned out to be a negative result when the data were analyzed properly.
4. Another aspect of data handling is the need to consider what is referred to as the power of the study. If a study has low power, it means that there were not enough subjects recruited to achieve a specified probability of detecting a specific magnitude of benefit, and as a result even if the treatment made a substantial difference the study would not have been able to detect it. Freiman and colleagues reviewed seventy-one negative trials reported in the medical literature and found that in fifty of them the sample size was so small that the chance of detecting a true 50 per cent improvement with therapy was less than 90 per cent (Freiman et al., 1978). Therefore, one must be cautious in concluding that a therapeutic technology is not effective until one is satisfied that enough patients were recruited to detect improvement with therapy if that improvement had indeed existed.
5. Inadequate sample size can lead to a false negative conclusion, but neglecting to consider both clinical importance and statistical significance can result in a false positive conclusion. If sample size is sufficiently large, trivial differences between treatment and control can reach statistical significance. For example, the only positive finding in a recent control trial of digoxin was a very small (but statistically significant) difference in cardiac function measured by echocardiagram (Flegg, Gottlieb, and Lakatia, 1982). Therefore, we must consider not only whether a treatment effect exists, but also if it is large enough to mandate a change in therapy.

References

Adams, R. 1958. "Internal Mammary Artery Ligation for Coronary Insufficiency, Evaluation." *New England Journal of Medicine* 258: 113-115.

Baxter, W.D., and Levine, R.S. 1969. "An Evaluation of Intermittant Positive Pressure Breathing in the Prevention of Post-Operative Pulmonary Complications." *Archives of Sursgery* 98: 795-798.

Beecher, H.K. 1955. "The Powerful Placebo." *Journal of the American Medical Association* 159: 1602-1606.

Benson, H., and McCalla, D.P. 1958. "Angina Pectoris and Placebo Effect." *New England Journal of Medicine* 17: 945-952.

Bonchek, L.I. 1979. "Are Randomized Trials Appropriate for Evaluating New Operations?" *New England Journal of Medicine* 301: 44-45.

Cass, Principal Investigators and Their Associates. 1983a. "Coronary Artery Surgery Study (CASS): A Randomized Trial of Coronary Artery Bypass Surgery". *Circulation* 68: 939-950.

Cass, Principal Investigators and Their Associates. 1983b. "Coronary Artery Surgery Study (CASS): A Randomized Trial of Coronary Artery Bypass Surgery – Quality of Life in Patients Randomly Assigned to Treatment Groups." *Circulation* 68: 951-960.

Chalmers, T.C. 1975. "Randomization of the First Patient." *Medical Clinicians of North America* 59: 1035-1038.

Chalmers, T.C. 1982. "Potpourri of RCT Topics." *Controlled Clinical Trials* 3: 285-298.

Chatterjee, K.; Parmley, W.W.; Massey, B.; Greenberg, B.; Werner, J.; Klausner, S.; Norman, A. 1976. "Oral Hydralazine Therapy for Chronic Refractory Heart Failure." *Circulation* 54: 879-883.

Cobb,. L.A.; Thomas, G.I.; Dillard, D.H.; et al., 1959. "An Evaluation of Internal Mammary Artery Ligation by a Double-Blind Technique." *New England Journal of Medicine* 260: 1115-1118.

Committee 1975. "Committee Reviewing Controlled Trials of Hypoglycemic Agents." *Journal of the American Medical Association* 231: 583-608.

Diamond, E.G.; Kittle, C.F.; Crockett, J.E. 1960. "Comparison of Internal Mammary Artery Ligation and Sham Operation for Angina Pectoris." *American Journal of Cardiology* 5: 484-486.

Dohli, S., and Gold, M.I. 1978. "Comparison of Two Methods of Post-Operative Respiratory Care." *Chest* 73: 592-595.

Ellis, L.G., et al. 1958. "Long-term Management of Patients with Coronary Artery Disease." *Circulation* 17: 945-952.

Fields, W.S.; Maslenikov, V.; Mayer, J.S., et al. 1970. "Joint Study of Extracranial Arterial Occlusion." *Journal of the American Medical Association* 2: 1993-2003.

Flegg, L.; Gottlieb, S.H.; Lakatta, E.G. 1982. "Is Digoxin Really Important in Compensated Heart Failure?" *American Journal of Medicine* 73: 244-250.

Franciosa, J.A.; Pierpont, G.; Cohn, J.N. 1977. "Hemodynamic Improvement after Oral Hydralazine in Left Ventricular Failure." *Annals of Internal Medicine* 86: 388-393.

Franciosa, J.A.; Weber, K.T.; Levine, E.B.; Kinasewitz, G.T.; Janicki, B.S.; West, J.; Henis, M.M.J.; Cohn, J.N. 1982. "Hydralazine in the Long-term Treatment of Chronic Heart Failure: Lack of Difference from Placebo." *American Heart Journal* 104: 587-594.

Freiman, J.A.; Chalmers, T.C.; Smith, H.; Jubler, R.R. 1978. "The Importance of Beta, Type II Error and Sample Size in the Design and Interpretation of the Randomized Controlled Trial: Survey of 71 'Negative Trials.' " *New England Journal of Medicine* 299: 690-694.

Gent, M. 1984. Personal Communication.

McNeil, B.J.; Weichselbaum, R.; Pauker, S.G. 1978. "The Fallacy of the Five-year Survival in Lung Cancer." *New England Journal of Medicine* 299: 1397-1401.

Royal College. 1981. "Disabling Chest Disease: Prevention and Care. A Report of the Royal College of Physicians by the College Committee on Thoracic Medicine." *Journal of College of Physicians of London* 15: 69-87.

Rubin, S.A.; Chatterjee, K.; Ports, T.A.; Galberg, H.J.; Brundich, B.H.; Parmley, W.W. 1979. "Influence of Short-term Oral Hydralazine Therapy on Exercise Hemodynamics in Patients with Severe Chronic Heart Failure." *American Journal of Cardiology* 44: 1183-1189.

Sacks, H.S.; Chalmers, R.C.; Smith, H. 1983. "Sensitivity and Specificity of Clinical Trials." *Archives of Internal Medicine* 143: 753-755.

Sinclair, J.S. 1966. "Prevention and Treatment of the Respiratory Distress Syndrome." *Pediatric Clinics of North America* 13: 711-730.

Spodick, D.H. 1982. "The Randomized Controlled Clinical Trial." *American Journal of Medicine* 73: 420-425.

Chapter 6:

Guidelines for Health Technology Assessment: Diagnostic Technologies

Gordon Guyatt, Michael Drummond, David Feeny,
R. Brian Haynes & Peter Tugwell

The development of diagnostic technologies can be divided into three phases: the developmental phase, the exploratory phase, and the establishment of accuracy.

The Developmental Phase

New diagnostic technologies generally develop as a result of advances in the basic sciences. The evolution of new modes of diagnostic imaging, including computerized tomography (CT), ultrasonography, and nuclear magnetic resonance (NMR), are among the most exciting developments in medicine over the past decade. These innovations followed directly on advances made in physics laboratories. The discovery of new substances that circulate in the blood proceeded from experimentation in biology and immunology, and the clinical application of the measurement of these substances was subsequently established.

The developmental phase of a diagnostic technology begins with basic science experimentation and the development of a prototype which is then tested to determine its performance integrity. Image devices are tested to see how well they reproduce objects of various density (referred to as phantoms); biological and immunological assays are tested for their reliability and accuracy in evaluating specimens in which the concentration of the substance of interest is known. Initial estimates of possible toxicity can be obtained during the developmental phase, such as specifications of the amount of radiation associated with imaging devices. When the relevant properties of the device are known, and virtually all the fine tuning that can go on in the laboratory is done, the developmental phase has been completed.

The Exploratory Phase: Establishing the Range of Possible Benefit

When the technology moves from the laboratory into the clinical setting it makes sense to apply it to patients with a large number of diverse conditions. The goal of this exercise is to delineate the possible uses of the technology. When evaluating CT or NMR scanning of the head, for example, one might do scans on patients ranging from the very young to the elderly, including a large number of normals, and patients with haemorrhagic and thrombotic strokes, tumours, arteriovenous malformations, demyelinating disease, and so on. Echocardiography, which provides an image of the heart by recording reflected sound waves, would be administered to patients with a wide variety of cardiac conditions.

Striking impressions often follow from these preliminary studies. The ease with which CT scanning identifies haemorrhage, and the difficulty in delineating lesions in the first two weeks following thrombotic stroke, would be discovered. The ability of NMR to differentiate grey and white matter, and its utility in demonstrating demyelinating lesions (such as those that occur in multiple sclerosis) would be evident. The ease with which M-mode echocardiography demonstrates pericardial effusions and the difficulties in accurately delineating the severity of valvular lesions of the heart would be obvious.

The only criterion for patient selection during this exploratory phase (besides informed consent) should be a good idea of what the true underlying condition really is. Any attempt to limit the spectrum of patients seen will prevent a full appreciation of the range of results possible with the new technology. Those interpreting the tests need not be blind to the underlying condition of the patients being examined. In fact, the greater the amount of clinical information, the better, for this phase of instrument development is a learning process in which unexpected correlations are discovered and interpretations refined.

There are dangers in terminating this exploratory phase, and proceeding to more rigorous evaluation, prematurely. On the one hand, fruitless experiments, in which ultimately negative results could have been predicted beforehand, may be undertaken. On the other hand, the discovery of potentially worthwhile applications may be delayed. These dangers may appear extremely implausible at present, for the development of new diagnostic technologies has to date been characterized by a profusion of this sort of exploration and a paucity of properly controlled studies. What is perhaps more dangerous is that certain elements of more advanced studies (such as establishing reliability, carefully defining the patient population, or blinding the test interpreters) may be incorporated in what is still essentially an exploratory study. This would have two deleterious consequences: the full potential of the early phase of technology evaluation would not be realized, and a spurious impression of the adequacy of exploratory studies in establishing test utility may be created.

Another way to think of this phase of technology evaluation is as a hypothesis-generating stage. One would start with a general sort of question: what happens in NMR scanning of the heart in patients with myocardial infarction? What happens to carcinoembryonic antigen levels (CEA – a tumour marker) in patients after removal of their colon carcinomas? One would then perform NMR on all the heart attack victims and obtain CEA levels on all the patients with bowel tumours going to surgery that one could and examine the results obtained. From such studies would arise hypotheses such as: the results of NMR scanning in the first day after a heart attack predicts the subsequent development of heart failure; persistent evaluation of CEA levels after surgery indicates residual tumour. These hypotheses would then be tested in more rigorously designed studies (rather than being reported as established facts, as is too often the case).

In addition to their clinical uses, diagnostic technologies may contribute to a better understanding of human physiology and mechanisms of disease. Measurement of levels of thyroid hormone in the blood may not only help to diagnose hypothyroidism or hyperthyroidism, but may aid understanding of the metabolic adaptations of the body during serious illness. Positron emission tomography (PET scanning) is likely to provide important information about tissue metabolism in health and diseases, irrespective of its diagnostic utility. While elucidation of disease mechanisms constitutes an important reason for developing and using diagnostic technologies, this goal is largely independent of clinical considerations, and by itself would lead to much more limited dissemination than is presently seen. For these reasons, we shall limit our discussion to the accuracy and benefit of technologies in the clinical context.

In the hierarchy of evaluative goals presented earlier, the first, establishing technological capability, is accomplished during the developmental stage of the new technology. Separate guidelines are required for each subsequent stage and will now be addressed.

Accuracy

For a diagnostic technology to be clinically useful it must reliably distinguish between diseased and non-diseased, quantify the severity of an illness or condition, or both. The pitfalls encountered in trying to determine the accuracy of diagnostic tests, and the best ways around them, have been most thoroughly studied for tests that try to determine the presence or absence of disease, rather than its extent or severity (Department, 1981; Ransohoff and Feinstein, 1978). Based on this previous work, our present guidelines are designed to apply, as much as possible, to both situations.

Framing an Appropriate Question

It makes little sense to ask, how good is the CT scan? Or even, how accurate is the CT scan? We need to know exactly what is meant by good or accurate: good for what, in whom, and compared with what?

For technologies designed to confirm or refute the presence of a disease, "how good is" generally translates into "what is the sensitivity (proportion of diseased patients with a positive test) and specificity (proportion of non-diseased patients with a negative test) of each level of test results?" Posing the question in this fashion implies the need for an independent assessment, or "gold standard" that will allow us to conclude whether the results of the technology do or do not correspond to the "truth." More sophisticated measures than sensitivity and specificity have been developed for describing the accuracy of diagnostic technologies that produce a continuous result rather than just a positive/ negative result. Receiver operating characteristics (ROC) curves permit choice of the test cut-off between diseased and non-diseased that maximizes sensitivity and specificity (Swets, 1979). For example, in patients with lung cancer the decision whether their lesion is operable is often based on the presence of spread of the tumour to lymph nodes in the mediastinum (the part of the chest between the lungs). CT scanning has been used to determine the size of the mediastinal lymph nodes, and thus to help detect whether the mediastinal lymph nodes are involved with tumour in patients with lung cancer (the bigger the nodes, the more likely they are infiltrated with tumour). One might determine the best use of the CT scan in deciding whether the tumour has spread to the mediastinum by setting the size cut-off at the one that diagnoses spread of the malignancy at 1, 1.5, and 2 centimetres and by observing the effect on sensitivity and specificity. One would then choose the cut-off that maximized the accuracy of the test.

Another use of ROC curves is to set the cut-off so that the test will have a certain specificity (for example, a false positive rate of 10 per cent) and to calculate the sensitivity of the test at that specificity. Formulating their data in this fashion, Swets and colleagues were able to show that CT scanning had a higher sensitivity than did brain scanning at all levels of specificity (Swets et al., 1979). Statistical methods that allow one to determine which of two ROC curves is "better" (determined by calculating the area under each curve), and whether the difference is statistically significant, are available (Hanley and McNeil, 1982).

An alternative approach is to develop what are called likelihood ratios (LR). LR allow estimates of the probability that disease is present at any level of diagnostic test result (Department, 1983). Using the CT and mediastinal tumour example, one could determine the likelihood that tumour is present given a pre-test probability (the estimate of the likelihood that tumour has spread to the mediastinum which one makes before applying the diagnostic test), and node sizes of 1, 1.5 and 2 centimetres. These approaches, though sophisticated, are all straightforward variants of sensitivity and specificity, based on a comparison between the test results and a gold standard.

For diagnostic technologies that assess levels of function or severity of disease (such as, measurement of cardiac output by angiography or radio-nuclide (RN) scanning, lung function by spirometry, or exercise capacity by treadmill or walking tests), a way of quantifying the relation between the test and the gold standard (that is, answering the question "how good is" with a number or set of numbers) is less clear. When the units of measurement differ (as when one compares maximum power output on a treadmill exercise test with the distance walked in a six-minute walk test) a simple correlation coefficient can describe the relation. If the units of measurement are the same (as when one compares measurement of cardiac ejection fraction by radionuclide angiography with the ejection fraction determined by the gold standard, angiography) the intraclass correlation coefficient is most appropriate as it takes into account both systematic and random error (Kramer and Feinstein, 1981). An alternative is to use a statistic called Kappa, which measures the degree of agreement beyond chance (Cohen, 1961). A modification of this statistic, weighted Kappa, allows one to consider not only the fact that the technologies disagreed with the gold standard, but the extent of the discrepancy (Cohen, 1968).

These methods have a common problem that one does not face when considering dichotomous outcomes. Values of sensitivity and specificity are intuitively meaningful, especially given that one can translate them directly into the probability that a disease is present. It is considerably less clear, however, what a simple correlation, or intraclass correlation, or weighted Kappa, of 0.5 actually means about the accuracy of a diagnostic technology. Still, these methods are the best available for answering the question "how good is" a technology in evaluating the extent of disease and will have to do for the present.

The next issue, the need to know in whom and for what, is illustrated by the potential use of CT scanning of the mediastinum in patients with lung cancer, still an important and controversial issue. At present, in most North American centres, patients with a diagnosis of (non-oat cell) carcinoma of the lung, whose tumour is larger than 2 or 3 centimetres on a chest X-ray and in whom no evidence of distant metastases is found on initial screening, undergo mediastinoscopy. In this minor operation, the lymph nodes of the mediastinum are explored: if the procedure is negative the surgeon procedes to thoractomy and lung resection for cure; if it is positive, the patient is given palliative therapy. It has been suspected that CT scanning of the mediastinum may eliminate the need for mediastinoscopy.

One could go about investigating the accuracy of CT scanning in this context in various ways. One might collect a heterogeneous group of patients with various thoracic lesions in whom surgery is planned and compare the results of CT scanning with operative and pathological findings, resulting in estimates of sensitivity and specificity for CT of the thorax. Alternatively, one could do the same sort of study in a heterogeneous group with suspected mediastinal pathology. In this case one could look at the prediction from CT for each of the chains of lymph nodes in the mediastinum and compare them with the surgical findings, once again arriving at figures for sensitivity and specificity. Neither of these two approaches would help the clinician who wonders about the need to keep subjecting patients to mediastinoscopy.

A better strategy for this purpose would be to do a CT scan on every patient in whom mediastinoscopy is planned. After all the surgery is complete, patients can be divided into two groups: those in whom resectable tumour was found (no mediastinal spread by either mediastinoscopy or subsequent thoractomy), and those in whom metastatic spread precluded resection for cure (tumour found on mediastinoscopy or, following negative mediastinoscopy, on thoractomy). The ability of the CT to distinguish between these groups could then be evaluated. This type of study would allow sensitivity and specificity estimates for a clinically meaningful outcome and would enable the clinician to decide about the further need for persisting with mediastinoscopy in lung cancer patients being considered for resectional surgery.

Framing an appropriate question also requires asking how good the new technology is compared with available alternatives. The study design for assessing a new technology's accuracy will depend on the context planned for its use. If the new procedure is intended to be used in addition to existing methods (echocardiography, for instance, is performed in addition to, rather than instead of, a chest X-ray) subjects should first receive the conventional work-up and an estimate of disease likelihood and severity should be made. The technology under investigation, and subsequently the gold standard, can be applied, and whether the new technology improved accuracy (that is, the combination of conventional work-up plus new technology more closely predicted the gold standard than the conventional work-up alone) can be determined. Conners, McCaffree, and Gray asked physicians to estimate pressures in the pulmonary circulation in critically ill patients in an intensive care unit, based on all available clinical information. Right heart catheterization served as the gold standard and did indeed provide additional information, in that physicians were unable to predict the findings of catheterization (Connors, McCaffree, and Gray, 1983).

That a technology's accuracy does not offer information relevant to diffusion until it is compared with existing methods is illustrated by studies of the use of CT scan in patients with headache (Larson, Omenn, and Lewis, 1980; Carrera et al., 1977). It was found that the CT scan has a sensitivity of 100 per cent in detecting identifiable organic disease in patients with headache. This is quite impressive and might suggest we screen headache patients using CT scanning; as it turns out, however, the sensitivity of a simple physical examination is also 100 per cent.

A new diagnostic technology may be designed to replace, rather than complement, existing methods. For example, CT head scanning has served as an alternative (at least in some patients) to the previously available imaging procedures such as radio-nuclide brain scanning, pneumoencephalography, and angiography. In this situation, a direct comparison between two or more technologies is appropriate. Swets and colleagues (1979) compared CT with RN scanning in 136 patients who had undergone both procedures. Their gold standard for the diseased patients was histological examination of tissue, and for the non-diseased, negative results of other tests and either a finding of no intracranial lesions at autopsy or live follow-up with no neurologic symptoms for at least eight months after the imaging. Scans were read separately with appropriate blinding of readers, and the results clearly demonstrated the superior accuracy of the CT scan.

There may be situations in which both technologies cannot be applied to all patients because of considerations of time, cumulative toxicity, or patient acceptance. In such situations, an equally valid, though less efficient procedure is to randomize patients to receive one technology or the other, with both groups being assessed by the gold standard. When the study is completed, direct comparisons of sensitivity and specificity, or chance-corrected agreement, can once again be made.

Another purpose of diagnostic technology assessment may be to delineate the factors that influence accuracy. In this situation one need not use actual patient material. For example, investigators have used human cadaver "phantoms" to delineate the effect of search strategies, light, distance, mind-set, film characteristics, and other features on interpretation of chest radiographs (Brogdom, Kelsey, and Mosely, 1983).

Comparison with a Gold Standard

Unless one compares a new diagnostic technology with an independently determined gold standard, there is no way to determine its accuracy. For this reason, comparison with a gold standard is usually regarded as a sine qua non for adequate evaluation of new diagnostic tests. The gold standard is often a more invasive or dangerous procedure (if it did not have some problem associated with it, there would be no need to seek an alternative), such as coronary artery angiography for electrocardiographic stress testing. In many situations, however, it would be unethical to go on to a definitive procedure. When looking at the accuracy of CT scanning in suspected brain tumour, for example, one would not undertake surgical exploration to confirm negative tests. In this situation long-term follow-up may serve as an adequate gold standard for a negative test: if a patient is well one year after a negative CT scan, he or she did not have a brain tumour at the time the scan was done.

In many situations this primary rule for determining accuracy, comparison with a gold standard, is not applicable—when no gold standard exists, for example. The histamine challenge test is designed to measure the severity of asthma. Patients inhale a substance that tends to make their airways constrict, and the concentration of the substance that is needed to produce a certain degree of bronchoconstriction is taken as an index of the severity of the underlying disease. Unfortunately, physicians have a difficult time defining asthma, let alone deciding on a gold standard for measuring its severity.

In this situation one must rely on construct validity. To demonstrate construct validity one examines the relation between a new test and existing measures and looks at whether the new technology relates to other variables in the way one would expect if it really is measuring what it is supposed to measure. Diurnal variation in airflow, bronchoconstriction in response to breathing cold air or exercising, and response to bronchodilator are all indices of asthma severity, and the fact that the histamine challenge test relates closely to all of them supports its status as a measure of asthma severity.

Precision and Observer Variation

For diagnostic technology to be useful it must produce more or less the same result each time it is performed on a person in the same clinical state. Similarly, if the test requires subjective interpretation, it must be established that readers can agree with themselves (intraobserver variation) and with one another (interobserver variation). Generally, when observer variation is studied systematically, it is surprising how much trouble even expert readers have agreeing with one another (and even with themselves). This difficulty exists even for seemingly objective procedures like venography, in which extensive retraining of experienced radiologists was required before observer agreement was increased to acceptable levels (McLachlan et al., 1979).

To avoid bias in comparing old and new technologies, every effort must be made to optimize the reliability of each procedure. For example, a study will not yield a fair comparison if it is conducted in a centre with a special interest and expertise in one technology and little interest or skill in a competitor. Investigators must, therefore, assure that the technical aspects of all test administration were optimized and that high-quality interpretation was available for all procedures.

Blinding

In establishing the extent of intraobserver and interobserver agreement, as well as the sensitivity and specificity of a technology with respect to a gold standard, each test must be conducted independently, without knowledge of other results. If this is not the case, the results of the investigations will, almost inevitably, influence one another, spuriously increasing precision and accuracy. Interpretation of the results of the technology under investigation will be influenced by knowledge of the "truth" obtained from the gold standard (such as when an equivocal stress test is interpreted in the light of a markedly positive angiogram). In addition, the intensity of the diagnostic work-up may be influenced by test results (Ransohoff and Feinstein, 1978). An extreme example of this phenomenon occurred in a study in which only patients with a positive stress test underwent angiography, and the patients with a negative stress test were assumed to be disease free (Goldschlager, Selzer, and Cohn, 1976). Another example of the consequences of neglecting blinding was the nitro-blue tetrazolium test for diagnosing infection. The test, which required subjective interpretation, enjoyed brief popularity as a result of apparently high sensitivity and specificity generated in unblinded studies. Unfortunately, the results could not be reproduced in studies in which blind interpretation was applied (Ransohoff and Feinstein, 1978).

While it is clear that interpretation of the technology being studied and the gold standard must be blind with respect to one another, the issue of what clinical information to provide the reader is more complex. One strategy, used by Swets in his study of CT versus brain scanning, is to provide readers with minimal diagnostic information (Swets gave only the patient's age and the technical features of the scan being read). That is certainly not what will happen in real life, however, when the test interpretation will be made with at least a brief clinical sketch such as "onset of left hemiparesis and severe headache six hours before test was conducted." Whether the presentation of this information to the test interpreter makes optimal use of the available investigations is a matter of dispute, but that is the way it happens. It is conceivable that the differences between CT and RN scanning that Swets observed would not have been present if the clinical information had been available. Let us assume that the CT scan provides images that lead to greater confidence in the interpretation of difficult problems in which the reader of the RN scan was forced to equivocate or guess. But what if the clinical information was such that it did not help the CT reader (who already had all the information needed to make the right choice) but provided the RN reader with data he or she needed to make the right choice in difficult cases? In this situation we would conclude that, when interpreted without clinical information, CT scanning is more accurate than RN scanning, but that in the clinical situation their accuracy is the same.

Whether there really are instances in which differences in accuracy are minimized or disappear (or are magnified) when clinical information is supplied remains to be determined. Whatever choice the investigator makes as to the amount of clinical information supplied to the test interpreter, the information should be identical, and blindness with respect to the gold standard result must be maintained.

Appropriateness of Patients Tested

It would be of little use to know that radio-nuclide angiography was capable of distinguishing between patients in pulmonary edema and healthy people coming in off the street: a glance at these individuals would be sufficient to make the distinction. We need to know whether diagnostic technologies are capable of identifying those with mild as well as severe disease, and distinguishing them not from normals, but from people with conditions easily confused with the disease of interest. These conditions strongly suggest that studies to determine accuracy be conducted by examining consecutive patients with a suspected condition, performing the diagnostic technology under investigation, and proceeding with independent application of the gold standard.

The level of angiotension converting enzyme in the blood has been proposed as a diagnostic test for sarcoidosis (a multisystem disease of unknown cause that most frequently affects the lungs in a manner somewhat similar to tuberculosis). It has been reported in numerous studies to have a sensitivity of approximately 60 per cent and a specificity of over approximately 90 per cent (Lieberman et al., 1979; Fanberg et al., 1976; Ashutosh and Keighley, 1976). The specificity has been determined in a wide variety of pulmonary conditions, however, most of which bear no clinical resemblance to sarcoidosis. In the single study in which controls were selected because of a suspicion of sarcoidosis, specificity was only 67 per cent, suggesting that the test would not be particularly useful in the clinical situation in which it would actually be needed (Schultz, Miller, and Bedrosian, 1972).

A second illustration of the need to choose an appropriate sample for determination of accuracy is illustrated by experience with exercise radio-nuclide ventriculography as a diagnostic test for coronary artery disease. When initially studied, radio-nuclide ventriculography had a specificity of 94 per cent; in subsequent use the specificity fell to 49 per cent (Rozanski et al., 1983). The explanation is that the disease-free population was far healthier during the earlier studies, and that patients who have a substantial pre-test likelihood of coronary artery disease (that is, the ones in whom we need the diagnosis sorted out) have a high incidence of non-specific abnormalities on exercise radio-nuclide angiography.

These examples emphasize that the accuracy of new diagnostic technologies must be established in the clinical context in which they will ultimately find their use.

Applicability

We have emphasized the importance of an accurate and detailed description of a therapeutic manoeuvre in allowing its application by others in their own clinical practices. This issue is equally relevant for diagnostic technologies. In addition to specifying the mechanics of performing the test, patient issues, such as special requirements for diet and activity, and precautions taken during the test should be clearly described.

Effect on the Health Care Provider

Accurate diagnostic tests may have no influence on either therapy or patient outcome and still receive wholehearted support from the medical community. CT scanning has been widely disseminated and received endorsement from a variety of official agencies and consensus conferences, without good scientific evidence that patients benefit from its application. This may be because physicians and policy-makers are convinced of the benefit for patients even in the absence of rigorous evidence because physicians are willing to argue for dissemination irrespective of benefit; or because the CT scan's ability to reduce the physician's anxiety and increase his or her confidence may be a powerful force favouring rapid diffusion and official endorsement. We suspect that this latter explanation may be the most important. Without doubt, it is immensely reassuring to know that a patient with an acute change in neurological status is not suffering from a condition (such as acute or chronic subdural hematoma) that requires immediate surgical intervention. Similar reassurance accompanies the knowledge that severe headaches of recent onset do not represent a brain tumour, or that the comatose patient with documented malignancy in whom one has decided to do nothing more does, in fact, have tumour deposits throughout the brain. Such reassurance is even more powerful in these days of rampant medical litigation in which a mistake may have disastrous consequences not only for the patient but for the physician as well.

Fineberg, Bauman, and Sosman (1977) did not include health professional reassurance in their hierarchy of evidence regarding diagnostic technology but did acknowledge this omission: "This efficacy hierarchy is purposely patient-oriented and ignores possible consequences of CT affecting only the clinician, such as the reassurance that everything is being done for the patient, satisfaction from employing the most advanced diagnostic tests, or personal financial gain."

Assuming that diagnostic technologies do have an effect on health workers, what part should this effect have on our decisions about resource allocation? If an expensive technology reassures the health care provider, but does not influence patient outcome, for example, should it be adopted? The effects of new technology on health workers should influence our decisions about the use of diagnostic technologies, but careful thought needs to be given to the appropriate magnitude of this influence.

Reasonable discussion of this so-far-neglected area cannot begin until presently unavailable data are collected about the extent to which diagnostic technologies provide reassurance and reduce anxiety for health workers.

Therapeutic Effect

A test result may have a diagnostic effect and still not affect therapy because a health worker may ignore the results of the test or be unaware of available treatment; there may be no therapy available; or the patient may have already been receiving the best possible therapy. If a diagnostic technology is to result in appreciable changes in morbidity or mortality, the information it provides must result in modification of therapy. If test results can be shown to lead to institution of an intervention of known effectiveness, patient benefit from the technology is certain. If unproven therapy changes, improvement in health status as a result of the diagnostic technology remains a possibility.

How would one go about showing that therapy has changed as a result of a diagnostic technology? By now the reader will probably be able to guess our response: a randomized clinical trial. In such a trial patients would be randomized to one of two treatment plans, only one of which would include the technology under investigation. The new investigation might be an add-on (such as adding exercise stress testing before leaving hospital to the investigation of the post-myocardial infarction patient), or the two arms of the experiment might contain alternative technologies (CT versus RN brain scan). In any case, the methodologic guidelines we have laid out for the investigation of therapeutic technologies are equally applicable to attempts to determine whether therapy changes as a result of a diagnostic test.

Like other types of control trials, those that examine the effect of a diagnostic technology on therapy are expensive and logistically difficult. One strategy that has been used to circumvent these problems is to ask clinicians what they would do before giving them a diagnostic result and then give them the results and see if their plan changes. There are situations comparable to insulin therapy for diabetic coma in which such before-after "physician judgement" designs convincingly demonstrate that therapy will change as a result of diagnostic information. If the results of a diagnostic technology are rapidly available to the clinician, and changes in therapy are made immediately on receipt of the information, a causal relation between the test and the modification of therapy can be confidently presumed. The study by Conners, McCaffree, and Gray (1983) of right heart catheterization in the intensive care unit is an example of such a situation, for the authors looked not only at whether physicians could predict catheterization results, but also at whether the information received led to changes in therapy. Before inserting the catheter physicians were asked to record a treatment that they would pursue if they were not going to have catheterization results available; after thirty of sixty-two catheterizations, drug therapy or volume replacement was altered. This study convincingly demonstrates that right heart catheterization alters therapy in critically ill patients who have not responded to a therapeutic trial.

Using physician judgement studies based on clinicians' reports of therapeutic plans is enticing for a number of reasons. First, the expense and logistical difficulties of a randomized trial are avoided. Second, no patient is denied a potentially beneficial technology: everyone receives the diagnostic test. The problems of before-after designs, described in the following five subsections, detract, however, from their usefulness and should be carefully considered before accepting the results of studies with this design.

The Research Question

Physician judgement studies are better suited for add-on rather than replacement technologies. As an example of such an add-on study, Wittenberg, Feinberg, Ferrucci and colleagues (1978) found that a change in therapy attributed to the investigation occurred in 14 per cent of patients undergoing a CT body scan. In this study, physicians referring their patients for scanning were asked to list possible diagnoses, planned investigations, and tentative therapy, before the scan was performed. At the time of discharge, physicians were re-interviewed and asked if the CT scan had helped or hindered the diagnostic process and whether it affected therapy. Patient charts were reviewed to ensure that the physician interview accurately depicted what had happened. In this study CT examination was performed only after other appropriate non-invasive imaging studies were completed, never as a screening technique. Presumably this means that patients with suspected pancreatic lesions, for example, were studied by ultrasound before CT scan.

It would have been much more difficult to use the design to compare ultrasound with CT scan, a clinical question that is also important. If one is going to compare technologies rather than study a single technology as an add-on procedure, the design of studies in which clinician reports of planned therapy are used to estimate therapeutic effect becomes considerably more complicated. If one were comparing CT with NMR head scanning (a situation in which few would argue that patients should be subjected to both procedures), one would either have to randomize patients to one or the other technique, and compare how physician plans changed before and after receiving results of the two procedures, or have to subject patients to both procedures, and ask physicians to describe how each result would modify treatment plans. If the first choice is made, one is back with an RCT with all its difficulty and expense, and the logistics of the second option appear formidable. It appears, therefore, that the physician judgement design is a feasible and appropriate approach only for add-on, rather than replacement, technologies.

Discrepancy between Stated Plans and Clinical Action

Clinicians' stated plans may differ from what the clinicians would actually have done had the diagnostic test not been available. Studies by both Wittenberg and colleagues (1978) and Robbins and colleagues (1978), for example, found that CT scans of the body apparently saved a substantial number of surgical procedures. In many instances the decision to perform surgery or other invasive interventions is a difficult one. In situations where the right decision is not apparent it may be much easier to say that surgery is the planned therapy when one knows that the opportunity to rescind the decision is still available. To say that one will operate in a borderline situation contingent on the CT result may be quite different from actually taking the patient to the operating room. The same may be true of other treatment plans that entail substantial risk to the patient.

Causation and Subjective Judgement

If one asks the physician ordering a test to decide whether the result had a therapeutic effect, that physician will have an investment in the test's proving useful. Evidence on the bias that this can create comes from a study by Goldman and colleagues (1980) who found that physicians ordering echocardiograms felt that 26 per cent of the studies provided information that resulted in a change in management. Two independent reviewers felt, however, that only 8 per cent of the echocardiograms were needed for a change in patient management. One solution to the problem of relying on the subjective judgement of the physician involved to determine therapeutic effect is to use independent reviewers. Unfortunately, the use of independent reviewers does not solve the problem. If the person making the judgement believes that the diagnostic technology is beneficial, or if they find that it provides them with reassuring information, they might like to see a positive result to the study. If this is so, conscious or unconscious bias may lead to generosity in attributing changes in therapy to the technology. Of course, a clinician poorly disposed to the technology could bias the study in the opposite direction. At any rate, these biases are likely to exist and are difficult to measure.

Ensuring Validity of Data

When physicians are asked to review the diagnosis and therapy that patients finally received, they may err. Fineberg addressed this problem by reviewing patient charts to ensure that physician interviews accurately depicted what had happened.

Conclusion

It may be that despite this host of problems, studies of therapeutic effect based on a before-after look at physicians' plans and ultimate treatment may produce valid results. The only way to know for sure is to do what has been done for uncontrolled trials or those using historical controls: compare them with RCTs asking the same question and see if the results are the same. If the answer is yes, the cheaper and less troublesome before-after studies will be the way to go. The only preliminary evidence we have in this regard is the observation that unnecessary surgery was avoided by CT, which both Fineberg and Robbins made in their investigations of the therapeutic effect of CT body scanning. In the only RCT of CT scanning (Kingham et al., 1981) in which patients presenting with undiagnosed abdominal masses were randomized to CT or conventional imaging, the proportion of the patients who went to laporotomy was actually higher (39 per cent) in the group receiving CT scans than in the controls (32 per cent). Although the sample size of this study was very small, and thus the power very low, the results reinforce our suspicion that when RCTs looking at what really happens are compared with before-after studies using physician plans, the result will be the same as when RCTs of treatments are compared with other studies of interior design: the latter studies lead to substantial overestimates of the benefit of the intervention. Only further work in this area, however, will prove whether we are right or wrong.

Methodological Standards

The criteria described in the following four sections are important considerations in deciding whether to implement a controlled trial or a before-after study and in deciding whether to accept the results.

Clinical Context of the Study

Studies should focus on a specific patient presentation or clinical problem. In the early phases of the clinical application of a diagnostic technology one sees case reports describing how the new tool was able to make a diagnosis that resulted in change in treatment, whereas existing methods were not. Such reports are now appearing on nuclear magnetic resonance versus CT. These reports are interesting, but not particularly helpful in deciding whether NMR does have a therapeutic effect. The situation reported may have been idiosyncratic, gives us no idea of the frequency with which NMR provides information that changes therapy, and does not tell us how often misleading results occur.

In many studies of therapeutic effect the protocols included patients with several different problems without providing the results for each problem; thus it is often difficult to understand the effect on a specific clinical presentation. It is considerably more useful to know that CT scanning changes therapy for, say, suspected pancreatic carcinoma, than to know that CT body scanning changes therapy. Most work to date has focused on the technique (CT head scanning, CT body scanning) rather than on the presenting problem. In their study of body scanning, Wittenberg, Fineberg, Ferrucci and colleagues (1978) had clinicians and investigators classify body CT scans as (1) providing information that changed diagnosis, prognosis, or therapy, (2) providing new information but without clinical effect, or (3) providing no new information. They found that chest CT scans were most useful for evaluating hilar and mediastinal masses. However, details of the presentation of these patients was not provided. In this study patients entered into one of ten protocols; some of the protocols, however, included several quite different problems. The lung protocol, for example, included the problems of hilar and mediastinal staging, as well as that of the peripheral lung nodule. The best way of obtaining an appropriate sample is to study consecutive patients presenting with a specific problem.

True Accuracy of a Test Whose Effect is Being Assessed

It is important that the accuracy of the test be established beforehand or concurrently with the demonstration of therapeutic effect as described in the section on accuracy. Even CT scanning of the head, one of the most accurate diagnostic technologies available, has a sensitivity of 92 per cent and a specificity of 90 per cent (Swets et al., 1979). This means that 8 per cent of diseased patients will have negative scans, and 10 per cent of those without disease will be labelled as having an important abnormality. It is possible, therefore, that CT can change therapy for the worse when patients are mislabelled. Indeed, 1 per cent of the patients in the study by Wittenberg and colleagues of body scanning were labelled by their physicians as having received therapy that was not in their best interests as a result of information from the CT scan. CT scanning also encouraged surgery that proved to be unnecessary in four patients. It would be reassuring to know that this problem was of no greater magnitude than the authors discovered. To do so, we would require gold standard evaluation of the sort that Swets demanded: pathological evidence of disease in the positive patients and asymptomatic status after long-term follow-up in the patients labelled disease-free.

Identification of Patients Most Likely to Benefit

There will be patients in whom the pretest likelihood of disease is so low that a positive result in the new test will not lead to institution of therapy (or so high that therapy would be administered despite a negative result). The information gain is maximal for patients with pretest probabilities of 30 to 70 per cent (Pauker and Kassirer, 1980). The threshold for prescribing therapy will depend on the seriousness of the condition as well as on the toxicity of the therapy. In addition, there is also wide variation in the threshold of the probability of disease that needs to be attained for different physicians to make the same specific therapeutic decision (Pauker and McNeil, 1981).

Therapeutic Effect and Patient Outcome

Even if a technology has a real therapeutic effect, it does not mean that its clinical usefulness is established. That is only true if the therapy instituted on the basis of the test does more good than harm (and/or the therapy curtailed does more harm than good). In their paper on CT head scanning, Fineberg, Bauman, and Sosman (1977) describe thirty-six patients in whom treatment was modified on the basis of the results of scanning. Of these, seven died and nine were treated by surgery or radiation therapy, and "in each instance the surgeon or radio-therapist stated that CT results influenced the approach to or extent of treatment." New (unspecified) treatment was started for fifteen patients; for seven, (unspecified) anticipated treatment was cancelled; and for six, treatment was abandoned because the patient's condition was judged hopeless. Even this relatively detailed description of the changes in therapy that followed CT scanning leaves us somewhat unsure whether changes in therapy provided benefit. It is crucial to take into account the strength of existing evidence that the therapy of interest does more good than harm. The following section describes whether assessment of patient outcome itself should be measured in studies of therapeutic effect.

Patient Outcome

As far as is possible, we should be convinced that patient outcome changes as a result of a diagnostic technology before it is disseminated. Outcome has often been narrowly defined in the past. In fact, patients may benefit from a diagnostic technology in many ways, including reduction in mortality, reduction in morbidity, and improvement in the quality of life. Reduction in morbidity should include decreases in any distressing symptoms or complications resulting from other more invasive tests. By "quality of life" we mean both a person's ability to undertake activities that he or she finds rewarding and enjoyable and the way he or she feels. A diagnostic technology may change quality of life even when other more easily measurable variables show no change. Sox, Margulis, and Sox (1981) found, for example, that of patients presenting with non-cardiac chest pain, those randomized to receive routine CPK measurements and electrocardiograms showed less short term disability than did the control group. Just as the physician may find that a negative CT decreases his or her anxiety about the possibility of a brain tumour in patients with the recent onset of severe headaches, so may the victims of the pain. This reassurance may be extremely important for the worried patient.

The therapeutic value of the diagnostic tests themselves, as the reassurance value might be labelled, is worthy of considerably more investigation than it has received, but we would like to include one caveat for those who might try. The reassurance value of the test for the patient may be confounded with the reassurance value for the physician, and unless the latter is measured, the extent of this confounding may be impossible to sort out. Patients with severe headaches who undergo a CT scan, for example, may have a stronger conviction that they do not have a brain tumour than their counterparts who are spared the test only because the physicians in the latter group expressed some hesitation about the matter. The conclusion from such a study would not be to recommend CT scans for every patient with a headache, but to educate physicians so that they realize that if a careful neurological examination is negative in such a patient it virtually rules out a brain tumour. They might then be able to provide the reassurance that would quell the patient's anxieties.

Why do we insist on patient benefit before concluding that a diagnostic technology has passed the clinical assessment and is ready for economic scrutiny? The issue may be clarified by review of the study by Conners, McCaffree, and Gray (1983) investigating right heart catheterization. (Other aspects of the clinical assessment did not allow physicians to predict pulmonary vascular pressures. Therapy was altered in 48 per cent of the patients on the basis of test results.) Is this data sufficient to recommend dissemination of the technology? (The question is somewhat academic since the technology is already disseminated, but useful for illustrating the point.) Fifty-seven per cent of the patients in the study died during hospitalization, so one can quickly conclude that the catheterization or subsequent changes in therapy did them little good. Could the technology have improved the outcome for at least some of the remaining 43 per cent? It is certainly possible, but the authors do not provide any data one way or the other. Right heart catheterization entails serious risks, including death from catheterization ventricular arrhythmias. Therefore, although Conners and colleagues have demonstrated the diagnostic effect of right heart catheterization, they have not shown that the procedure does more good than harm. Even before examining economic aspects (and the technology is indeed expensive), one would require evidence of benefit before considering dissemination of the technology.

Can one cite instances in which a diagnostic technology provided information but failed to change clinically relevant outcomes? In one case, application of a diagnostic test (serum cholesterol), when followed by a specific therapy (clofibrate), actually increased mortality (Committee, 1980). Emergency endoscopy in patients with upper gastrointestinal bleeding provides increased diagnostic information without altering rates of surgery, hospital stay, or mortality (Drofield et al., 1982; Peterson et al., 1981). Chest radiographic screening for lung cancer fails to change outcome (Brett, 1969). These examples illustrate the wisdom of demonstrating improvement in patient outcome before allowing the dissemination of diagnostic technologies.

The decision whether to carry out a controlled trial on therapeutic management that includes assessment of patient outcomes depends to a large extent both on feasibility and on the existence of previous RCTs demonstrating effectiveness of the therapy that is influenced by the diagnostic technology in question. If controlled trials demonstrate that application of a diagnostic technology leads to the institution of a therapy proven effective by previous RCTs or to the termination of harmful therapy (as might happen when a patient without a disease is mistakenly given a toxic treatment), benefit can be considered established. An example is impedance plethysmography (IPG), a test for the diagnosis of deep venous thrombosis (leg clots), one of the case studies to which we shall apply our guidelines (see Chapter 9). Because the anticoagulant heparin is known to do more good than harm in deep venous thrombosis, that IPG leads to appropriate administration of heparin is a sufficient demonstration of its usefulness. Even in these studies, however, the authors were forced to look directly at patient outcome to ensure that no harm followed from withholding heparin from those whose IPG was negative. Given the difficulty of performing studies in which differences in outcome are demonstrated, it may be worthwhile to concentrate diagnostic technology assessment in areas in which existing RCTs have shown that treatment does more good than harm.

When Are RCTs Measuring Patient Outcome Unnecessary?

As with RCTs of therapeutic technologies, the most stringent tests of benefit are not appropriate in some situations. These include: (1) Patient benefit from the test is so dramatic that even observational studies leave no room for doubt. Electrocardiography for dysrhythmias associated with treatment of known effectiveness is one example. CT head scanning for head trauma is sufficiently dramatic in decreasing the need for exploratory surgery that it can probably also be included in this category (Zimmerman et al., 1978). (2) The new technology is equally or less toxic and equally or less expensive than existing alternatives and has been shown to be more accurate. However, this is still a criterion only for bypassing the establishment of benefit in an RCT if the technology that is being replaced is either itself of known benefit or is here to stay despite the absence of demonstrable benefit. Pneumoencephalography, for example, was never shown to improve outcome in patients with neurological disease, but in the absence of planned trials of the benefit of the procedure it was appropriate to replace it with the equally accurate, or more accurate, and much less harmful CT scan.

Conduct of Studies Examining Effect on Patient Outcome

If it is decided to implement a study to examine the effect of a diagnostic technology upon patient outcomes, the design should be a randomized control trial. The reader may be wearying of our blandishments regarding the need for RCTs, but once again the guidelines we suggested for therapeutic technologies apply to establishing the effect of diagnostic technologies on patient outcome. Patients must be randomized to receive or not to receive the diagnostic test (or to receive one of two diagnostic tests if the new technology is intended to replace existing methods) and then followed over time to see if one group or the other does better.

If the reader still has any doubt about the need for concurrent randomized controls, an additional example directly relevant to diagnostic technologies may help to settle the issue. A group of Australian neurologists examined the effect of CT scanning on mortality in stroke patients by comparing the outcome of stroke patients seen in 1978 who were investigated with CT with that of patients seen in 1974 before the scan was introduced (Christie, 1979). To ensure that it was the CT scan that was making the difference, they chose patients matched for all the prognostic variables they thought relevant. The 1978 group had a lower mortality, apparently providing dramatic evidence of the effect of CT on outcome. They also, however, examined mortality in another matched group: stroke patients seen in 1978 when the CT scan was not working. This group had the lowest mortality of all, comparable to the other 1978 patients and lower than the 1974 group.

The conclusion is that such before-after studies are inadequate for assessing the effect of diagnostic technologies on patient outcomes. In this instance the reason for the false impression of improvement of patient outcome is probably that changes in the management of stroke may have occurred (such as better monitoring, blood pressure control, nursing) that do not require or are not affected by the CT scan result. It is also possible that the patients in the more recent group were not as sick in ways that the investigators failed to measure (as were the historical control group). This type of error leads us to believe that concurrent randomized controls are almost always necessary to establish the benefit of diagnostic technologies when effects on patient outcomes are to be assessed.

To conclude our guidelines for the clinical assessment of diagnostic technologies, we will make several points about the special challenges of RCTs of diagnostic technologies. First, RCTs examining diagnostic technologies will be strengthened if diagnostic possibilities and therapeutic plans are elicited from health care providers before and after application of the technology, for this will help to clarify the mechanism of any effects that are found.

Second, patient selection for studies of diagnostic technology must be appropriate. As disucssed in the section on therapeutic effect, for some patients the pre-test likelihood of disease will be so low that even a positive result on the new test will not lead to institution of therapy (or so high that therapy would be administered despite a negative test).

Third, the true accuracy of the test should be established, because one would not expect benefit from an inaccurate test. In studies where the decision is made to proceed directly from the preliminary stage of establishing the range of possible use to RCTs in which patient benefit is the primary measure of outcome, mechanisms for assessing the accuracy of the technology can be built into the trial. If this approach is taken, the hierarchy of accuracy, effect on the health care provider, therapeutic effect, and patient benefit can all be examined in a single study.

Fourth, all relevant outcomes, including complications of diagnostic testing, morbidity, mortality, and quality of life, must be measured.

References

Ashutosh, K., and Keighley, J.H.F. 1976. "Diagnostic Value of Serum Angiotensin – Converting Enzyme Activity in Lung Diseases." *Thorax* 31: 552-557.

Brett, C.Z. 1969. "Earlier Diagnosis and Survival in Lung Cancer."*British Medical Journal* 4: 260-262.

Brogdom, B.G.; Kelsey, C.A.; Masely, R.D. 1983. "Factors Affecting Perception of Pulmonary Lesions." *Radiologic Clinics of North America* 21: 633-654.

Carrera, G.F.; Gerson, D.E.; Schnur, J.; McNeil, B.J. 1977. "Computed Tomography of the Brain in Patients with Headache or Temporal Lobe Epilepsy: Findings and Cost Effectiveness." *Journal Computer Assisted Tomography* 1: 200-203.

Christie, D. 1979. "Before-After Comparisons: A Cautionary Tale." *British Medical Journal* 2: 1629-1630.

Cohen, J.A. 1961. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurment* 20: 37-46.

Cohen, J.A. 1968. "Weighted Kappa: Nominal Scale Agreement with Provision for Scale Disagreement or Partial Credit." *Psychological Bulletin* 70: 213-220.

Committee of Principal Investigators, World Health Organization. 1980. "Co-operative Trial on Primary Prevention of Ischemic Heart Disease Using Clofibrate to Lower Serum Cholesterol: Mortality Follow-up." *Lancet* 2: 379-385.

Connors, A.F.; McCaffree, D.R.; Gray, B.A. 1983. "Evaluation of Right Heart Catheterization in the Critically Ill Patient Without Acute Myocardial Infarction." *New England Journal of Medicine* 308: 263-267.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1981. "How to Read Clinical Journals: II To Learn about a Diagnostic Test." *Canadian Medical Association Journal* 124: 703-710.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1983. "Interpretation of Diagnostic Data: V How to Do It with Simple Maths." *Canadian Medical Association Journal* 129: 22-29.

Drofield, M.W.; Langman, M.J.S.; Atkinson, M.; et al. 1982. "Outcome of Endoscopy in Barium Radiography for Acute Upper Gastrointestinal Bleeding: Controlled Trial in 1,037 Patients." *British Medical Journal* 284: 545-548.

Fanberg, B.L.; Schoenberger, M.D.; Bachus, B.; et al. 1976. "Elevated serum Angiotensin One Converting Enzyme in Sarcoidosis." *American Review of Respiratory Diseases* 114: 525-528.

Fineberg, H.B.; Bauman, R.; Sosman, M. 1977. "Computerized Cranial Tomography." *Journal of the American Medical Association* 238: 224-227.

Goldman, L.; Feinstein, A.; Batsford, W.P.; et al. 1980. "Ordering Patterns and Clinical Impact of Cardiovascular Nuclear Medicine Procedures." *Circulation* 62: 680-687.

Goldschlager, N.; Selzer, A.; Cohn, K. 1976. "Treadmill Stress Tests as Indicators of Presence and Severity of Coronary Artery Disease." *Annals of Internal Medicine* 85: 329-335.

Hanley, J.A., and McNeil, B.J. 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." *Radiology* 143: 29-36.

Kingham, J.G.C.; Fry, K.K.; et al. 1981. "Computed Tomography in Patients with an Abdominal Mass: Effective and Efficient?" *Lancet* 1: 1199-1201.

Kramer, M.S.; and Feinstein, A.R. 1981. "Clinical Biostatistics LIV: The Biostatistics of Concordance," *Clinical Pharmacology Therapeutics* 26: 111-123.

Larson, E.B.; Omenn, G.S.; Lewis, H. 1980. "Diagnostic Evaluation of Headache." *Journal of the American Medical Association* 243: 359-362.

Lieberman, J.; Nosal, A.; Schlesner, L.A.; et al. 1979. "Serum Angiotensin – Converting Enzyme for Diagnosis and Therapeutic Evaluation of Sarcoidosis." *American Review of Respiratory Diseases* 120: 329-335.

McLachlan, M.S.F.; Thompson, J.F.; Taylor, D.W.; et al. 1979. "Observer Variation in the Interpretation of Lower Venograms." *American Journal of Roetenology* 132: 227-229.

Pauker, S.G.; and Kassirer, J.P. 1980. "The Threshold Approach to Clinical Decision Making." *New England Journal of Medicine* 304: 1109-1117.

Pauker, S.G., and McNeil, B.J. 1981. "Impact of Patient Preferences on the Selection of Therapy." *Journal of Chronic Disease* 34: 77-86.

Peterson, W.L.; Barnett, C.C.; Smith, H.J.; et al. 1981. "Routine Early Endoscopy in Upper Gastrointestinal Tract Bleeding. A Randomized Trial." *New England Journal of Medicine* 304: 925-929.

Ransohoff, D.F.; and Feinstein, A.R. 1978. "Problems of Spectrum and Bias in Evaluating the Efficacy of Diagnostic Tests." *New England Journal of Medicine* 299: 926-929.

Rozanski, A.; Diamond, G.A.; Berman, D.; et al. 1983. "The Climbing Specificity of Exercise Radionuclide Ventriculography." *New England Journal of Medicine* 309: 518-522.

Robbins, A.H.; Pugatch, R.B.; et al. 1978. "Observations on the Medical Efficacy of Computed Tomography of the Chest and Abdomen." *American Journal of Roetenology* 131: 15-19.

Schultz, T.; Miller, W.C.; Bedrósian, C.W.M. 1972. "Clinical Application of Measurement of Angiotensin – Converting Enzyme Level." *Journal of the American Medical Association* 242: 439-441.

Sox, H.V.; Margulis, I.; Sox, C.H. 1981. "Psychologically Mediated Effects of Diagnostic Test." *Annals of Internal Medicine* 95: 680-685.

Swets, J.A. 1979. "ROC Analysis Applied to the Evaluation of Medical Imaging Techniques." *Investigative Radiology* 14: 109-121.

Swets, J.A.; Pickett, R.M.; Whitehead, S.F.; et al. 1979. "Assessment of Diagnostic Technologies." *Science* 205: 753-759.

Wittenberg, J.; Fineberg, H.V.; Ferrucci, J.T.; et al. 1978. "Clinical Efficacy of Computed Body Tomography II." *American Journal of Roetenology* 131: 5-14.

Zimmerman, R.A.; Bilaniuk, L.T.; Gennarelli, T.; et al. 1978. "Cranial Computed Tomography in Diagnosis and Management of Acute Head Trauma." *American Journal of Roetenology* 131: 27-34.

Chapter 7:

Guidelines for Health Technology Assessment: Economic Evaluation

---

Michael Drummond

Chapters 5 and 6 were principally concerned with assessing whether a new health care technology (therapeutic or diagnostic) does more good than harm when applied in a clinical setting. Chapter 7 presents guidelines that attempt to build on the clinical evaluation by assessing whether the benefits of the new technology exceed the benefits forgone (that is, the costs) by applying the same resources to their best alternative use. In fact, it is this *opportunity cost*, or sacrifice, implicit in the rapid growth of technology that should be the community's real concern, not the dollar amounts themselves. The dollar amounts are only a measuring rod for the opportunity costs.

Economic evaluation has two objectives: first, to introduce resource considerations into the analysis and to assess the opportunity costs of new technologies; second, to develop a framework within which the costs of the new technology can be compared with its benefits. The following section introduces methodological guidelines for undertaking an economic evaluation of new health care technologies.

Choosing a Service or System Approach

A major design decision for the economic analyst is whether to evaluate the new health care technology from a service or a system perspective. In the system, or "macro" study the perspective is that of the entire health care system. In the service, or "micro", study the focus is more narrow. Consider the evaluation of cimetidine, an H_2-receptor antagonist effective in the treatment of peptic ulcer disease. A service approach to economic evaluation would be closely integrated with the clinical evaluation of this drug. The objective would be to undertake economic analysis in association with any controlled clinical trials being carried out, for example, cimetidine versus surgery (such as proximal vagotomy). In this way analysts can ascertain the relative costs of the new technology versus the old. (The different levels of sophistication in making this comparison are discussed later.)

Alternatively, a system approach to economic evaluation could be adopted, where the objective is to estimate the total economic effect (defined as of costs and benefits) of the new technology. For cimetidine, for example, a system approach would involve an analysis of the changes in health service costs of the new technology. Some costs may result from inappropriate uses of cimetidine, for instance in clinical situations where it has not been shown to do more good than harm or where there is an equally effective, but cheaper, alternative (for example, antacids or cessation of smoking to the pain like that of ulcers). Studies of many new health care technologies show that use is not confined to situations where effectiveness or cost-effectiveness has been demonstrated (Larson et al., 1981; Bodemar et al., 1980.)

Of course, the two approaches to economic evaluation described above are complementary to each other and are not necessarily alternatives. If the guidelines for clinical evaluation described earlier in this volume were strictly adhered to, only the service approach would be needed in that technologies would only be appropriately applied. Because this ideal is hardly ever achieved, however, the system approach can alert us to the opportunity costs of the new technology and may serve as a stimulus for re-examination of its uses.

Most of the methodological points discussed below apply equally well to both approaches to economic evaluation. Most of the discussion, however, will be about service studies, as these are more closely integrated with the methods of clinical evaluation discussed earlier.

Identifying the Viewpoint for the Analysis

The resources for health care typically come from several budgets. In Canada the provincial governments incur most of the costs via the health insurance plans, which pay for physicians' services, and global budgets to hospitals. Government expenditures are not restricted to the Ministry of Health; other ministries may also provide services, especially for the elderly, for handicapped children, and for disabled adults. The patient and family also incur costs in obtaining health care. These are not restricted to financial outlays (in travel expenses or user charges) but include time costs in travelling to, waiting for, and receiving treatment and in providing informal home care for sick relatives. Finally, the community at large may bear part of the cost of treatment if it loses the output of individuals while they are receiving care.

Conversely, health care can confer benefits to all of these parties. If an individual is cured, or the onset of illness prevented, government health care expenditures and lost production may be reduced or averted. Also, the patient and his family may feel better in addition to saving money on family health care.

Economists typically argue that one should adopt a societal viewpoint for the evaluation of a new health care technology. This implies consideration of a wide range of costs and benefits in the economic evaluation. It would be false to conclude, for example, that a new technology such as day-care surgery is efficient if the analysis of comparative costs were confined to the hospital; studies have shown that there are often extra costs, such as community support services (Waller et al., 1978).

Analysis of the costs and benefits of new technologies by each viewpoint is important, however, if one wants to understand more fully their diffusion (see Chapter 3) or to develop options for health care policy (see Chapter 11). One of the reasons more family physicians have not adopted nurse practitioners, for example, could be that they could suffer a financial loss (Spitzer, 1978). Yet, new diagnostic imaging technologies, such as computed tomography (CT) scanning, that enable physicians to undertake examinations more quickly without in-patient admission may be financially attractive to providers regardless of the overall balance of costs and benefits to society.

A good economic evaluation of new health technology would therefore be made primarily from the societal viewpoint, although analysts may also need to consider other viewpoints to identify the net gains and losses to key participants in the health care system.

Considering a Relevant Range of Alternatives

Economics is about choice, given the scarcity of resources. Economic evaluations of new technologies must therefore include a thorough comparison of alternative courses of action. It is possible to find examples in the literature where only one alternative is assessed. In a study of a five-bed coronary care unit (CCU), Reynell and Reynell (1972) calculate the cost per life saved in the unit and suggest that "this kind of outcome analysis by survival could be used to assess the performance of other types of coronary care units and other forms of expensive medical treatment." Although it is important to find out whether coronary care units save lives and to ascertain their operating costs, the "assessment of performance" of expensive medical treatments is best carried out by comparison with possible alternative therapies. In this case the correct approach would have been to undertake a controlled study of the CCU versus relevant alternatives, such as an ordinary hospital ward and treatment at home, and then to compare costs and effectiveness.

A relevant range of alternatives may also not be considered if the "do nothing" alternative is omitted. Most of the economic evaluations of CT scanning, for example, have compared the costs and benefits of the new technology with existing diagnostic approaches (Thomson, 1977; Jonsson and Marke, 1977; Bartlett et al., 1978). Although such an approach is valid up to a point, it assumes that existing technologies have themselves been found to be effective and efficient when compared with "doing nothing". It will be clear from the discussion of the guidelines for clinical evaluation that this may not be a correct assumption. Nevertheless, one might still argue that it is worthwhile (in a second-best sense) to adopt a new technology that accomplishes a given therapeutic or diagnostic objective at lower cost, even if it has not been established that either technology does more good than harm. This would, however, be a poor second-best.

Another way in which a new technology may not be appropriately compared with an alternative is where an ideal form of the new procedure is compared with the existing technology in current use. Costs for conventional radiography, for example, cover the numerous retakes (about 10 per cent of all exposures) necessary because of inadequacies in patient preparation, X-ray procedure, or film processing. In theory, digital diagnostic imaging almost eliminates the need for hard copy on film and would therefore lead to cost savings in technician time, film, and darkroom facilities (Drummond et al., 1983). (With digital diagnostic imaging, a video image is instantly available.) In practice, however, it may be that for medical or legal reasons or because of problems in clinician and radiologist acceptance of reporting from video images, more hard copy will be made than was originally intended (although it still may be less than with conventional radiography). One can think of other examples, as in surgery, where not every surgeon is as competent as the pioneers of a new procedure, and in drug therapy, where every physician may not have the skills to select the appropriate patients for therapy, to optimize dosage, or to gain the patient's co-operation with treatment (Shapiro et al., 1983).

Part of the answer to this problem lies, of course, in the conduct of the clinical trials themselves. It was mentioned earlier in this volume that there ought to be more management trials where the new technology is compared with the old in a practical setting.

Selecting the Appropriate Form of Economic Evaluation

One of the objectives of economic evaluation is to develop a framework within which the costs of a new technology can be compared with its benefits, and different levels of sophistication exist in making the comparison. It is possible to distinguish five forms of economic evaluation: cost analysis, cost-minimization analysis, cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis.

All five forms consider costs and differ primarily in their assessment of benefits. In evaluating a new technology it is important to select a form of analysis consistent with the question to be answered. Unless something were already known about the relative benefits of the new and existing technologies, an analysis of costs alone would not suffice. The following examples will illustrate the importance of selecting the appropriate form of economic evaluation. (For a more detailed description of the different forms of analysis see Department of Clinical Epidemiology and Biostatistics, 1984.)

Although most forms of economic appraisal consider both costs and benefits, a simple *cost analysis* of the new technology, compared with existing treatment options, may occasionally suffice. This true case in the economic assessment of the oxygen concentrator, a new electrically driven machine that extracts oxygen from air (Lowson, Drummond, and Bishop, 1981). The new machine was a potential replacement for oxygen cylinders in the provision of long-term domiciliary oxygen (up to fifteen hours a day) for chronic bronchitics. The effectiveness of the different methods of oxygen production could be assumed to be the same in this particular case, so interest centered on the relative costs. A major difference between the concentrator and the other methods was that it required capital outlay for purchase of the machines and provision of workshop facilities. The analysis of costs showed, however, that for all but small numbers of patients for whom the set-up costs per patient would be prohibitively high, concentrators were preferrable on economic grounds.

A slightly more complex situation is one in which the effectiveness of the alternative therapies is not known, but can be shown to be equivalent by a controlled clinical trial. The studies by Russell and colleagues (1977) and by Waller and colleagues (1978) day-care and short-stay surgery, respectively, are good examples of *cost-minimization analyses*. In both cases a randomized controlled trial was conducted to assess whether the new technology was as effective as traditional in-patient care, measuring by the number and types of complications, length of convalescence, and recurrence rates. Because the effects were found to be approximately the same, the economic analysis turned on the comparative costs.

In the last example the two therapeutic alternatives were equivalent in medical terms; but what should be done if the clinical effects differ? In an early *cost-effectiveness analysis*, Klarman, Francis, and Rosenthal (1968) argued that treatments for chronic renal failure have one major objective, the extension of life. If this were true, it would be legitimate to compare options by their cost per life-year gained. (The logic here, of course, is that with a limited budget to be devoted to treatment of a given disease, selection of options with the lowest cost per life-year gained would maximize the total extension of life.) Although now updated by the work of other researchers (such as Ludbrook, 1981), the work by Klarman, Francis, and Rosenthal remains a good example of the kind of results that can be obtained without directly facing the question of benefit valuation.

The next logical consideration would be the *quality* of the life-years gained. In the case above, the therapeutic alternative that might be assumed to give the higher quality of life (transplantation) was already the preferred option because of its simple cost-effectiveness. Nevertheless, this may not generally be true and other therapeutic alternatives that embody a conscious trade-off between length of life and quality of life, such as chemotherapy for some types of cancer may also exist. Clearly, if economic appraisal is to assist in such situations, it needs to address the quality of life.

A study that uses a quality-of-life approach to benefit measurement is that by Stason and Weinstein (1977). Their purpose was to undertake an economic evaluation of the management of essential hypertension, both to determine how resources could be used most efficiently within programs to treat hypertension and to provide a yardstick for comparison with alternative health-rated uses of these resources. The principal issue in measuring benefits was that blood pressure sceening and control, though possibly extending life, also has unpleasant side effects. Also, extension of life is not the only potential benefit of lowered blood pressure; non-fatal (but disabling) strokes and myocardial infarctions may also be averted. A *cost-utility analysis* was therefore performed and the results expressed as cost per quality-adjusted life year gained. (In utility analysis the preferences of members of the public for different states of health are measured. This technique will be explained more fully below.)

As in the study by Klarman, Francis, and Rosenthal (1968), the adjustment chosen by Stason and Weinstein was not based on thorough empirical analysis, but at least the authors did explore the sensitivity of their results to the assumptions made. There are, however, studies in which researchers rigorously obtained empirical utility measures. Some of these will be mentioned later. In addition, the case study in Chapter 10 on neonatal intensive care employs this approach.

Cost-utility analysis begins to address the issue of *valuation* of improved health by questioning whether the improvement in health brought about by the new technology is worth having, in relation to the resources used. Cost-minimization analysis, by contrast, tacitly assumes that the treatment objective is worth meeting. The expression of the results of cost-utility analysis as cost per quality-adjusted life year gained does not of itself answer the question posed above, but the comparison of these indices across a broad range of health treatments or programs has been suggested as a way of forming health care priorities (Kaplan and Bush, 1982).

Finally, researchers sometimes express as many as possible of the costs and benefits of options in dollars. This form of analysis, known as *cost-benefit analysis*, was adopted by Weisbrod, Test, and Stein (1980) in their economic assessment of a new community-oriented program for mentally ill patients compared with a "traditional" hospital-oriented program. In theory, a cost-benefit analysis ought to be able to answer the question, is the new technology worthwhile? In practice, however, few studies provide estimates in dollars of all the costs and benefits, and many intangible items are omitted. Therefore, if cost-benefit analysis is restricted to a consideration of easily measureable costs and benefits, it is often a narrower form of analysis than cost-utility analysis. Nevertheless, these studies can provide useful information. The contribution of the study by Weisbrod, Test, and Stein were estimates of the costs to the various agencies involved, to patients and families, and to the community at large.

In the early stages of the development of a new health care technology it may be difficult to predict which form of economic analysis is required. In any case the choice of design may often be an *ex post* rather than *ex ante* decision; it may be that in some situations the original intention is to conduct a cost-minimization analysis but that, contrary to expectations, the two therapies give rise to different effects. Nevertheless, the original conception of the new health care technology, in addition to the early clinical experiments, may give the economic analyst a few clues. New health care technologies tend to be one of two types: those that attempt to provide a new capability and those that attempt to replace existing procedures by a less costly, or more beneficial, method. The evaluation of the former type is likely to require a more sophisticated form of analysis, because the benefits of having the new capability (in diagnosis or therapy) need to be assessed and compared with the costs. For some technologies of the latter type, a simpler form or analysis may suffice.

Identifying the Relevant Range of Costs and Benefits

A crucial determinant of the relevant range of costs and benefits is the viewpoint for the economic analysis. It was argued earlier that a societal viewpoint should be adopted. Published studies vary in the extent to which they investigate the whole range; apart from the adoption (either explicitly or implicitly) of a more restricted viewpoint, there are two other common reasons for omissions.

First, capital costs are often omitted, perhaps because they are often charged to a different budget and hence recorded in a different place in health-service accounts. Omissions do not often occur in obvious cases. We do not know of any evaluations of CT scanning, for example, that omit the capital cost of the scanner; but it is quite common to find that the capital costs of the workshops required or the extra rooms in which the machine and its computer are to be located are not considered. (It should be remembered that space in hospital buildings does have an opportunity cost if it is being denied other uses.) On the benefit side, if a new health care technology frees space within the hospital, this may avert capital expenditure in the future.

Second, costs and benefits that, for one reason or another, are difficult to measure are often omitted. On the cost side, items that have a readily available price in dollars, such as supplies and personnel time, are always included; but those that do not, such as patients' time and volunteer time, are often omitted. We shall say more about the methods of valuing these items later, but the principle is clear: the diversion of resources from other beneficial uses implies an opportunity cost. Similarly, on the benefit side, the intangible aspects of improved health are often not adequately considered in economic analyses, measured benefits often being restricted to the more narrowly defined economic benefits of averted lost production and savings in health service resources. Researchers are making progress, however, in the area of benefit measurement.

Whether the exclusion of relevant costs and benefits is crucial to the analysis will depend on the question for study and the extent to which analysts believe that unmeasurerd costs and benefits would merely confirm the existing study result. If, for example, digital diagnostic imaging (DDI) were to produce images as accurate as those from conventional imaging but at lower cost, it may be possible to make a decision to adopt the new technology without formally assessing the benefits from the lower invasiveness of DDI. Yet a decision about adopting a new technology for screening the unborn fetus to provide the option of aborting those with serious congenital anomalies might be very difficult to make without assessment of the psychic costs of abortion and the value to potential parents of avoiding the birth of a child with Down syndrome (Hagard and Carter, 1976).

Finally, several items of cost and benefit, although not generally ignored in economic evaluations, may be particularly important in the assessment of new health care technologies. A major item is likely to be the costs of change to the new technology, such as re-training and/or redeployment of staff, closure or change of use of health service buildings, provision of information on the new technology to patients or personnel, and the administrative costs of all these tasks. In addition, if the new technology requires new machinery, the costs for maintenance and repair may be important. (An associated issue, the length of life of new medical equipment, will be discussed later in this chapter.)

Costs to be included in an economic evaluation of a health care technology or program are of three types (Stoddart, 1982). The first is organizing and operating costs within the health care sector. These include the costs of supplies; of capital; of space, light, heating, and power; of the time of health care professionals; and of maintenance and operation of equipment. A second type of costs is that borne by patients and their families, including out-of-pocket expenses, time lost from work, psychic costs, and patient and family inputs into treatment. The third type is cost borne externally to the health care sector and to patients and their families, such as the costs of complying with safety regulations, of meeting health monitoring requirements, and of performing toxicology studies.

Estimating Costs and Benefits

As mentioned earlier, items that have an easily measurable price in dollars are rarely omitted in economic evaluations of new health care technologies. The source of valuation for these items is therefore usually the market price. If the evaluation is being carried out from a societal viewpoint, the tacit assumption is that the prices observed reflect the social opportunity costs of the use of resources. (This assumption has been the subject of debate in the economics literature; see, for instance, Sugden and Williams, 1978.) Normally we would expect economic evaluations to use market prices for resources whenever these exist, unless there is a compelling reason to do otherwise. Where market prices are clearly distorted, however, it may be necessary to undertake more complicated and detailed studies to impute a price free of distortion. In these cases, the sensitivity of the final results to the imputation will need to be carefully explored. Additional analyses may be necessary, for example, in evaluations that use price observations from markets in which price controls are in effect or in valuations of time for groups who clearly face discrimination in the labour market. A variety of tools are available for imputing prices that more accurately reflect the social opportunity cost of the resource.

Three commonly used market-based cost items deserve more comment. First, hospital per diem rates (or average operating costs per day) are often used to estimate hospitalization costs. This is quite important in the evaluation of new health care technologies. Some, such as coronary artery bypass surgery, may involve significant amounts of hospitalization; others, such as CT scanning, may reduce the need for patients to be admitted to hospital for diagnosis. There are a number of dangers in using the per diem rate to value the costs or benefits. First, it poorly reflects the different resource consumption by different categories of patients. Also, since a considerable proportion of hospital costs is not affected by patient volume, per diem rates are a poor guide to the *extra* savings or *extra* costs arising from treating slightly fewer or slightly more patients. We shall return later to this distinction between average and marginal costs.

Second, physician billings are frequently used to estimate medical costs. The methods of calculating rates of physician reimbursement differ from one health insurance plan to another. It is important to be clear about whether the different rates for procedures really reflect the differential amounts of time and expertise required and whether the rates include an element of overhead and other general expenses.

Third, market wage rates of patients are sometimes used to estimate the costs to society of patients' time taken up in treatment or the benefits, in averted production losses, from returning patients to good health and avoiding hospitalization.

Again, this practice has been the subject of debate in the economics literature (see Drummond, 1981). Patients' earnings should be used in economic evaluations of new health care technologies only when one has considered the following two questions: are we sure that production will be lost, and are we not allocating a higher priority in the access to care by high-income earners by including potential production losses in the economic analysis model?

If the valuation of those costs and benefits that have market prices sometimes presents difficulties, what about those for which there are no prices? Many of these items have already been mentioned, namely, volunteer time, patients' time (if not work time), and hospital or clinic space. One approach is to impute values by reference to market rates. Office rentals or hospital construction costs (per square meter), for example, could be used to value hospital or clinic space; wage rates for casual work might be used to value volunteer or patients' leisure time. Another approach, which might make more sense in a practical setting, would be to identify the services that may be foregone if extra clinic space is taken up or more volunteers used. In either case the important point is that such items are not to be lost to the analysis just because they do not have an easily identifiable cost, as measured by market price.

The major category of costs and benefits for which there is no easily obtainable valuation is that of the benefits of improved health other than any production gains or health-service resource savings. Early economic evaluations simply labelled these benefits as intangibles (Weisbrod, 1961). In recent years, however, researchers have made considerable progress both in estimating these in dollars and in valuing states of health relative to one another.

Acton (1975) and Jones-Lee (1976) have devised money estimates of the willingness to pay for improved health. Also, economists have investigated the implicit valuations of decision makers as mirrored by public investments in measures to promote health and to improve safety (Hurst and Mooney, 1983). The main contribution of this work has not been to arrive at a "right" valuation of health benefits, but to point out that items such as the extension of life are continually being valued and that more explicitness in this process would be beneficial. If nothing else, inconsistencies could be identified.

The alternative approach, of valuing health states relative to one another, is explored at length in one of the case studies in Chapter 10. In this approach, states of health are first described and then placed in order by a judge. (The identity of the judges has varied from study to study. Some studies have used physicians; others have used patients, policy makers, or members of the general public.) Following the ordinal ranking, the relative valuations are obtained by one of several methods. The time trade-off method, which has a certain amount of practical appeal, is to ask respondents how long they would be willing to spend in one state as opposed to another. Another approach, known as the standard gamble approach, has theoretical superiority. Respondents are asked to consider a gamble between two alternatives. Alternative A is a treatment with two possible outcomes: either the patient is returned to health and lives out his or her life (t years, the probability of which is designated p), or the patient dies immediately (probability, $1\text{-}p$). Alternative B has the certain outcome of a chronic state (i) for life (t years). Probability p is varied until the respondent is indifferent between the two alternatives, at which point the required preference value for health state i is simply p.

In recent years there has been considerable testing of the reproducibility of the various methods of obtaining preferences of health states as well as assessment of the agreement between approaches and the extent to which the preferences of doctors, patients, policy makers, and the general public differ (Wolfson et al., 1982; Torrance, 1982). Reliable measurement methods are available. Those who feel uncomfortable with placing dollar values on health benefits (as in cost-benefit analysis) regard cost-utility analysis as a more promising approach.

Distinguishing Between Average and Marginal Costs and Benefits

One example of the importance of distinguishing between average and marginal costs and benefits has already been given, namely, the use of the per diem rate to estimate savings in hospitalization costs. If a new technology cuts hospital inpatient stays for a given procedure from, say, seven to three days, there is no guarantee that four-sevenths of the resources will be saved; some costs will be fairly fixed regardless of the number of patients in the hospital and therefore cannot be avoided. Also it may not be easy to redeploy existing resources to the benefit of other patients, as any physician will testify who has tried to find a bed for one of his or her own patients while spare capacity exists elsewhere in the hospital. (In general the ability to redeploy resources depends on the time scale.)

The distinction between average and marginal costs and benefits is particularly important in the economic evaluation of new health care technologies because of the tendency of the debate surrounding them to polarize. Some believe expensive health care technologies, such as scanners, are a major cause of the health service's problems; others believe they can do no wrong. The truth is likely to lie somewhere between. Few new technologies that are widely accepted are likely to have *no* appropriate uses; it is a question of how often they are used and for what purposes. In economists' language, these decisions should turn on marginal costs and benefits. Consider the following examples.

In the early 1970s the American Cancer Society proposed a protocol of six sequential stool quaiac tests for the detection of asymptomatic colonic cancer. The test was highly sensitive, detecting 91.67 per cent of the cancers being sought in the screened population. Thus, the first test will detect 91.67 per cent of cancers; the second test, 91.67 per cent of the 8.33 per cent remaining to be found after the first test, and so on. The *average* cost per case detected in six tests on the same population was a modest $2,451, but the *marginal* cost per case detected by the sixth in the series (having already tested the same population five times) was more than $47 million (Neuhauser and Lewicki, 1975). Most examples in everyday clinical practice are likely to be less extreme, but the same principle applies. In an evaluation of computed tomography of the brain in patients with headache, for example, Larson, Omenn, and Lewis (1980) caculated that the cost of finding a brain tumour was around $1,265 for patients with abnormalities on neurological examination but $11,901 or more for patients with no such findings.

Similar examples can be found in the use of therapeutic technologies. Neutra (1977) demonstrated, in a study of the indications for the surgical treatment of suspected acute appendicitis, that to save one more life by changing the symptom severity used as an indication for operation, one would need to perform so many operations on normal patients that the marginal cost of saving one more life would be $43 million and 2,053 person-years of convalescence. One would expect any economic evaluation of health care technology to recognize, therefore, the distinction between average and marginal costs and benefits.

Allowing for Differential Timing of Costs and Benefits

The time profile of costs and benefits may often differ between the new and the old technology. A surgical operation for duodenal ulcer, for example, will require that the majority of the costs be incurred now, with minimal costs in the future, providing the operation is successful. By contrast, medical treatment with cimetidine implies a steady stream of costs stretching into the future. Similarly, on the benefit side, a program to combat coronary heart disease placing a considerable emphasis on curative measures may give rise to benefits sooner than one with a preventive emphasis.

It is generally argued that the community is not indifferent to the timing of costs and benefits; we prefer to postpone costs but receive benefits sooner rather than later. The most widely accepted method of incorporating this notion into economic evaluations is to discount costs and benefits occurring in the future to present values. The details of the discounting calculation are given in Drummond (1980), along with more discussion of the underlying logic. The main point to note here is that the effect of the procedure is to give less weight in the analysis to costs and benefits occurring in the future.

A common way this issue arises in the evaluation of health care technologies is when the adoption of the new procedure requires a major capital purchase the first year, the rest of the expenditures being annual recurring amounts. Take, for example, an economic evaluation of the CT scanner. Thomson (1977) argued that the purchase price of the scanner, $259,994 (that is, 140,000 pounds sterling, at the 1977 pound sterling to Canadian dollar exchange rate), could be represented as $25,999 per year over its useful life. He then went on to compare this amount, plus the operating costs of the scanner, with the savings from the discontinuation of existing diagnostic procedures. If one applies the logic of discounting, however, an extra charge should reflect the capital costs incurred the first year. (This is similar to interest charged on a bank loan used to purchase a house or an automobile.) Using a discount rate of 5 per cent (in real terms) – the equivalent annual cost would be $33,669.

An even more important influence on the equivalent annual cost in this case is the assumption made about length of life of the equipment. Although one might not doubt that mechanically a scanner could last for ten years, its useful clinical life might be much shorter; clinicians may become dissatisfied with what they regard as "outmoded" equipment, or the hospital may feel that it is at greater risk of legal action if it is using equipment of lower technical capability than the newest machine on the market. If one were to assume a five-year life and a 5 per cent discount rate, the equivalent annual cost of the capital outlay would be $60,059 instead of $33,669. The useful clinical life of medical technologies is obviously important in determining their economic attractiveness, and therefore it would be prudent to make a fairly conservative estimate of this factor in any economic evaluation.

Allowing for Uncertainty in Costs and Benefits

The expected useful clinical life of a new technology is likely to be very uncertain at the time of the economic evaluation. Some major sources of uncertainty, such as the new technology's ability to improve patients' health status, may have been removed by the earlier clinical evaluation (provided that guidelines discussed in earlier chapters have been followed), although some of the longer-term outcomes, such as the reduction in radiation-induced tumours due to lower X-ray doses of the new digital techniques, may still be unknown. Other economic factors, such as the volume of patients served by new methods as applied in a practical setting, and the ability to redeploy freed resources to other beneficial uses, may be uncertain until the new technology has been used for some time.

One option would be to wait until there is enough experience with the new technology to obtain better estimates of all the variables in the economic model, as we have argued in the case of clinical uncertainties. This option is not, however, likely to be feasible. Once a new therapy proves to be effective, the pressure for its adoption is likely to be great, even if its efficiency is unproven; that is, even if a new technology is effective, the same resources may still generate more benefit if used in other programs.

Another strategy would be to make best estimates of the likely range of values of the uncertain parameters and then to undertake a *sensitivity analysis*. In this approach, the effect on the economic analysis of variation in the key parameters is explored. Stason and Weinstein (1977) have used this approach with some success in their study of hypertension therapy, as have Buxton and West (1975) in a study of renal dialysis. (This approach is also featured in the case studies in Chapter 9 and 10.) In fact, in the absence of reliable medical evidence for some new technologies, economists have applied sensitivity analysis to a number of the medical parameters. This can be a powerful approach if it can be shown that a new technology is not likely to be worthwhile, given its current costs, even if the effectiveness achieved exceeded the most optimistic medical predictions (Neuhauser, 1978).

More generally, the technique of clinical *decision analysis* provides a useful approach to analyzing the outcomes of a complex sequence of clinical alternatives (Weinstein and Fineberg, 1980). The effectiveness and efficiency of a new technology can be compared with the existing one by making informed estimates of the probabilities of the medical outcomes. We would still argue, however, that it is better to undertake an empirical assessment of the effectiveness of new technologies, if this is logistically possible and not prohibitively expensive.

Finally, it might be worthwhile to consider the net gains to key participants in the health care system that would accompany the introduction of a new technology under a variety of policy options. Although the economic evaluation is more properly done from a societal point of view, losses and gains to providers, insurers, and consumers often vary with the mode of introduction of the new technology, and the analysis can be augmented to identify these potentially differing patterns.

Presenting Results Appropriately

The final stage in an economic evaluation is the presentation of the results to those who will make decisions about the adoption of the new technology. It should be noted that there are decision makers at a number of levels in the health care system – for example, government policy makers, health service planners, and individual clinicians. Because the economic study results constitute the final link in our proposed evaluation scheme, it is important that the results are presented in a way that helps the decision maker appreciate their relevance to his or her own circumstances. In particular, evaluators should take note of the following three points when presenting their results.

First, it is important to identify *differences between locations*. For example, the type of clinical workload may vary, making the adoption of new technology more economically attractive in some locations than in others. Jonsson (1978) found this to be the case in Sweden for coronary care units and CT scanners. Also, pressures for hospital beds may differ among locations. This is particularly important in estimating the economic impact of a new technology in shortening hospital inpatient stays or in enabling more procedures to be carried out on an outpatient basis. If beds are in short supply, those freed may be used for the treatment of other patients, thereby reducing waiting times for admission (but increasing expenditures). If there is no pressure for beds, space capacity may be closed or extra provision averted in the future.

Locations may also differ in the resources already in existence – both physical facilities and clinical expertise – and in the volume of cases. Both of these factors could affect the costs per patient of adopting the new technology.

Many of these issues will be explored further in Chapter 8, on facilities planning, but it would also be beneficial if economic evaluations incorporated consideration of them as part of evaluation methodology. Some analysts have used sensitivity analysis fairly constructively in this regard (Jonsson and Marke, 1977; McGregor and Pelletier, 1978; Lowson, Drummond, and Bishop, 1981).

Second, it is important to identify key *value judgements*. It should be clear from the above discussion of economic evaluation methodology that a number of such judgements are involved. These include the choice of viewpoint for the analysis and the methods of measuring and valuing the costs and benefits. Evaluators should make these judgements explicit when results are presented and should conduct a sensitivity analysis to assess the effect of key judgements on the results of the study. Indeed, one of the chief strengths of a good economic evaluation would be that these judgements are made more explicit than they would be in alternative ways of arriving at a decision, such as "professional judgement." (Professional judgement and the current informal system of evaluation involve just as many value judgements as the formal economic analysis, but merely hide the sinister fact that they *are* being made.)

Finally, it is important to recognize that there are other relevant factors in decision making. Some of the better economic evaluations will discuss these when results are presented. A major factor is equity in the provision of services by geographical area or income class. Decision makers may sometimes make a conscious trade-off between equity and efficiency. A policy for location of, say, high technology diagnostic and therapeutic facilities based on equalizing patient travelling times may not be coterminous with one based on effective and efficient utilization of those facilities. The economic evaluation should estimate the cost, in terms of extra resources that could be redeployed elsewhere of such an equal access policy.

References

Acton, J.P. 1975. *Measuring the Social Impact of Heart and Circulatory Disease Programs: Preliminary Framework and Estimates*. Report R-1967-NHL1. Santa Monica, Calif.: Rand Corporation.

Bartlett, J.R.; Neil-Dwyer, G.; Banham, J.M.M.; Cruickshank, D.G. 1978. "Evaluating Cost-Effectiveness of Diagnostic Equipment: The Brain Scanner Case." *British Medical Journal* 2: 815-820.

Bodemar, G.; Gotthard, R.; Stom, M.; Walan, A.; Jonsson, B.; Bjurulf, P. 1980. "Socioeconomic Aspects of Treatment with Cimetidine in Peptic Ulcer Disease." In *Proceedings of a European Symposium on H2 Receptor Antagonists*. eds. A. Torsoli, P.E. Lucchelli, R.W. Brimblecombe. Amsterdam: Excerpta Medica.

Buxton, M.J., and West, R.R. 1975. "Cost-Benefit Analysis of Long-term Hemodialysis for Chronic Renal Failure." *British Medical Journal* 2: 376-379.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1984. "How to Read Clinical Journals. VII: To Understand Economic Evaluation (Part A)." *Canadian Medical Association Journal* 130: 1428-1434.

Drummond, M.F. 1980. *Principles of Economic Appraisal in Health Care*. Oxford: Oxford Medical Publications.

Drummond, M.F. 1981. "Welfare Economics and Cost Benefit Analysis in Health Care." *Scottish Journal of Political Economy* 28: 125-145.

Drummond, M.F.; Cockshott, W.P.; Haynes, R.B.; Walter, S. 1983. *Guidelines for the Evaluation of Digital Diagnostic Imaging Equipment or Units*. QSEP Research Report No. 86. Hamilton, Ontario: Faculty of Social Sciences, McMaster University.

Hagard, S., and Carter, F.A. 1976. "Preventing the Birth of Infants with Down's Syndrome: A Cost-Benefit Analysis." *British Medical Journal* 1: 753-756.

Hurst, J.W., and Mooney, G.H. 1983. "Implicit Values in Administrative Decision Making." In *Proceedings of a Workshop on Health Indicators*, ed. A.J. Culyer. London: Martin Robertson.

Jones-Lee, M.W. 1976. *The Value of Life: An Economic Analysis*. London: Martin Robertson.

Jonsson, E. 1978. *Studies in Health Economics*. Stockholm: Economic Research Institute, Stockholm School of Economics.

Jonsson, E., and Marke, L. 1977. "CAT Scanners: The Swedish Experience." *Health Care Management Review* 2: 37-53.

Kaplan, R.M., and Bush, J.W. 1982. "Health-Related Quality of Life Measurement for Evaluation Research and Policy Analysis." *Health Psychology* 1(1): 61-80.

Klarman, H.E.; Francis, J.O.; Rosenthal, G.D. 1968. "Cost Effectiveness and Analysis Applied to the Treatment of Chronic Renal Disease." *Medical Care* 6: 48-54.

Larson, E.B.; Mack, L.A.; Watts, B.; Cromwell, L.D. 1981. "Computed Tomography in Patients with Psychiatric Illnesses: Advantages of a Rule in Approach." *Annals of Internal Medicine* 95: 360-364.

Larson, E.B.; Omenn, G.S.; Lewis, H. 1980. "Diagnostic Evaluation of Headache: Impact of Computerized Tomography and Cost-Effectiveness." *Journal of the American Medical Association* 243(4): 359-362.

Lowson, K.V.; Drummond, M.F.; Bishop, J.M. 1981. "Costing New Services: Long-term Domiciliary Oxygen Therapy." *Lancet* 1: 1146-1149.

Ludbrook, A. 1981. "A Cost-Effectiveness Analysis of the Treatment of Chronic Renal Failure." *Applied Economics* 13: 337-350.

McGregor, M., and Pelletier, G. 1978. "Planning Specialized Health Facilities: Size vs. Cost and Effectiveness in Heart Surgery." *New England Journal of Medicine* 299(4): 179-181.

Neuhauser, D. 1978. "Cost Effective Clinical Decision Making: Are Routine Pediatric Preoperative Chest X-rays Worth It?" *Annales de Radiologie* 21: 80-83.

Neuhauser, D., and Lewicki, A.M. 1975. "What Do We Gain from the Sixth Stool Guaiac?" *New England Journal of Medicine* 293(5): 226-228.

Neutra, R. 1977. "Indications for the Surgical Treatment of Suspected Appendicitis: A Cost-Effective Approach." In *Costs, Risks, and Benefits of Surgery*, eds. H.P. Bunker, B.A. Barnes, and F. Mosteller. New York: Oxford University Press.

Reynell, P.C., and Reynell, M.C. 1977. "The Cost-Benefit Analysis of a Coronary Care Unit." *British Heart Journal* 34: 897-900.

Russell, I.T.; Devlin, H.B.; Glass, N.J.; Newell, D.J. 1977. "Day-Case Surgery for Hernias and Haemorrhoids: A Clinical, Social and Economic Evaluation." *Lancet* 1: 844-847.

Shapiro, M.; Schoenbaum, S.C.; Tager, I.B.; Munoz, A.; Polk, F. 1983. "Benefit-Cost Analysis of Antimicrobial Prophylaxis in Abdominal and Vaginal Hysterectomy." *Journal of the American Medical Association* 249(10): 1290-1294.

Spitzer, W.O. 1978. "Evidence that Justifies the Introduction of New Health Professionals." In *The Professions and Public Policy*, eds. P. Slayton and M.J. Trebilcock. Toronto: University of Toronto Press.

Stason, W.B., and Weinstein, M.C. 1977. "Allocation of Resources to Manage Hypertension." *New England Journal of Medicine* 296: 732-737.

Stoddart, G.L. 1982. "Economic Evaluation Methods and Health Policy." *Evaluation and Health Professionals* 5: 393-414.

Sugden, R., and Williams, A. 1978. *The Principles of Practical Cost-Benefit Analysis*. Oxford: Oxford University Press.

Thomson, J.L.G. 1977. "Cost Effectiveness of an EMI Brain Scanner." *Health Trends* 9: 16-19.

Torrance, G.W. 1982. "Preferences for Health States: A Review of Measurement Methods." In *Clinical and Economic Evaluation of Perinatal Programs*, ed. J.C. Sinclair, Mead Johnson Symposium on Perinatal and Developmental

Medicine, Number 20. Symposium held in Vail, Colorado, June 6 - June 10, 1982.

Waller J.; Alder, M.; Creese, A.; Throne, S. 1978. *Early Discharge from Hospital for Patients with Hernia or Varicose Veins*. London: Department of Health and Social Security, HMSO.

Weinstein, M.C., and Fineberg, H.V. 1980. *Clinical Decision Analysis*. Philadelphia: W.B. Saunders and Co.

Weisbrod, B.A. 1961. *Economics of Public Health*. Philadelphia: University of Pennsylvania Press.

Weisbrod, B.A.; Test, M.A.; Stein, L.I. 1980. "Alternative to Mental Hospital Treatment: Economic Cost-Benefit Analysis." *Archives of General Psychiatry* 1: 400-405.

Wolfson, A.D.; Sinclair, A.J.; Bombardier, C.; McGeer, A. 1982. "Preference Measurements for Functional Status in Stroke Patients: Interrater and Intertechnique Comparisons." In *Values and Long-Term Care*, eds. R.L. Kane and R.A. Kane. Lexington, Mass.: D.C. Heath and Co.

Chapter 8:

Planning for the Provision and Utilization of New Health Care Technologies

Roberta Labelle

Introduction

The introduction of a new health care technology or the proposed expansion of existing technologies is inevitably accompanied by a series of questions and concerns about the efficacy, effectiveness, and efficiency of the technology. Assessment of these characteristics of the technology is, of course, imperative before its adoption (or rejection). But once these assessments have been made and evaluation results disseminated, a final set of questions still confronts health policy makers and planners: how many of the new devices, interventions, or services should be provided, and where should they be located? This chapter addresses the question of how to determine, and then distribute, the optimal quantity of a health care technology for a given population. Throughout, it assumes that the type of assessments described in Chapters 4 through 7 have been conducted according to the standards outlined and that the results indicate adoption of the technology. The following discussion will therefore examine various methods available to ensure that the diffusion of an effective technology proceeds along a rational and well-conceived path.

"Optimal Facilities Planning" Defined

Planners in all disciplines have attempted to assess systematically the optimal number and distribution of facilities, yet nowhere is the discussion as extensive and as rich as in the literatures of management science and economics. Consequently, in analyzing the problem of the spatial allocation of health care technologies, this chapter will draw heavily on a concept originally introduced in location theory known as "optimal facilities planning." Broadly defined, optimal facilities planning (OFP) is a process undertaken by planners and policy makers to ensure that the current and future size, number, and location of facilities maximize total benefits to the community.[1] For present purposes, facility means any diagnostic or therapeutic procedure, device, or service that falls into the broad category of a "new technology".

The techniques available for OFP range from the simplest descriptive analyses to complex mathematical simulation models. The appropriateness of each technique will be determined by the nature of the planning problem. Only those methods relevant to the analysis of the diffusion of health care technology will be considered in this chapter.

Why is Planning for the Optimal Number and Location of New Technologies Important?

In 1976, in Southern California there were more computed tomographic scanners than in all of the United Kingdom (Shapiro and Wyman, 1976). A survey conducted by the American Hospital Association in 1969 indicated that about 35 per cent of institutions doing cardiac surgery perform fewer than one operation per month, and 70 per cent perform fewer than one operation per week (Platt, 1971). In Chicago, patients must travel on the average eight times farther for hospital care and twelve times farther for clinical care than they would if access to the nearest facility were available to them (Nelson, 1976). It is difficult to defend these scenarios on the basis of conscious, rational planning efforts; moreover, they are not unique. Blatant examples of under- and over-capacity are commonplace, and the implications for the efficiency and equity of the health care delivery system are profound.

In a society in which resources are scarce and wants seemingly unlimited, the introduction or expansion of a new technology necessarily implies forgoing expenditure on the next best alternative, be it in the health sector or any other sector. Thus, the total cost of purchasing or producing a health care technology must be balanced (at the very minimum) against the potential benefit from alternative technologies that are sacrificed in order to attain it.

Inadequate or excessive provision of a beneficial technology, as well as its inappropriate distribution, constitutes a pattern of resource allocation that, from a societal point of view, is inefficient, inequitable, or both. It is therefore of paramount importance to consider both size and distribution in the analysis of a new technology.

The Concept of an Optimum

OFP applied to the diffusion of new technologies requires addressing three basic questions. (1) How many of the new facilities should be provided? (2) What size should they be?

(3) Where should they be located? The questions are often highly interrelated and consequently cannot be answered independently of one another. As well, all are not equally relevant for every technology. When planning the diffusion of CT scanners, for example, size is a less important issue, but it is a very important consideration in planning a coronary care unit (CCU).

To answer the above questions, it is first necessary to establish specific criteria for determining an "optimum." The criteria employed in this chapter are the attainment of equity and efficiency in the provision of health care services. Accordingly, the next section proceeds by providing a definition of equity and efficiency. Methodological options and data requirements for determining the optimal *number* (or total provision) of facilities are then presented, first assuming that equity is employed as the guiding criterion, and second assuming that efficiency is the guiding criterion. The same format is then used in subsequent sections to examine the issues of optimal *size* and optimal *location* of individual facilities.

The Optimal Number of Facilities

This section addresses the question of how many new facilities should be purchased or provided, given the current state of knowledge regarding the technology. The question is as relevant for the introduction of new technologies as it is for the expansion of existing ones.

Determination of the Optimal Number of Facilities Using Equity as an Objective

Equity Defined. No consensus exists on the interpretation of equity in health care. Moreover, because "equity sometimes appears to be all things to all men" (Mooney, 1983, p. 179), the task of establishing a set of criteria to assess equity becomes extremely difficult. Fortunately, as Mooney (1983, page 179) notes, researchers generally agree upon two concepts fundamental to the notion of equity: "However equity is defined, all definitions contain some view of fairness of the distribution of something or other. It is also related in some way to the idea of a fair distribution across different individuals and/or groups in society." These groups may frequently be distinguished by income; other characteristics, however, such as age, sex, race, ethnicity, marital status, or geographical location, are equally applicable. Presumably in a health care system such as Canada's, with comprehensive, universal public insurance, barriers to access and utilization because of the above characteristics have largely been removed, and thus the more important concern is the equitable provision of health care based on the needs of the population.

This definition of equity, however, still presents difficult practical problems. First, several concepts of need exist in the literature,[2] and although many authors have grappled with the problem of securing an operational definition of need, a consensus on what actually constitutes need remains elusive. For the purposes of this chapter, and for consistency with its use throughout the book, "need" is defined as the level of provision indicated on the basis of an epidemiologically determined assessment of the requirements of the population. Need is restricted to situations in which there is an effective health care technology. Therefore, as defined here, need is a static rather than a dynamic concept, although in reality the needs of the population change as new knowledge is gained both on the nature of disease and on the diagnostic and therapeutic capabilities of technologies.

A second problem with the broad definition of equity presented above is that fairness in distribution is a vague concept, difficult to put into effect. By an equitable distribution do we in fact want provision sufficient to ensure equal access, equal utilization, equal quality of care or equal health?

There is no straightforward answer to this question, largely because the definition of equity is in part determined by the planning task at hand. Thus this chapter will offer three definitions of equity to correspond with the three stages of optimal facilities planning. Equity in the quality of care is best discussed within the concept of the optimal *size* of the facility. Geographical equity of access is an issue for the *location and distribution* of facilities. Both of these concepts will be addressed in subsequent sections. It is first necessary to determine the optimal *number* of facilities that will ensure equity of access across all classes of need.

Equity of access across all classes of need requires that sufficient numbers of the new technology are provided to ensure that no unmet need exists, or that unmet need is distributed randomly in the population. In addition, new technologies must not be supplied beyond the point indicated by the needs of the population; that is, there should be no excess capacity resulting from over-provision of the technology. These conditions are necessary, but not sufficient, for utilization to reflect need.

Need, however, is not an absolute concept. In a society of scarce resources, provision must be constrained by the available funds. Needs must therefore be ranked and priorities assigned to ensure that the correct amount of a given technology is provided *relative* to the provision of all other technologies. That is, the *mix* of health care technology must be optimal. (This condition is fulfilled when the marginal social benefit derived from the last dollar of expenditure on any given technology is equal across *all* technologies.)

Therefore, when examining whether the provision of CT scanners, for example, is equitable, health care planners must address two questions. First, is the current number of scanners sufficient to allow all those in need of services of the technology access to them? Second, are relatively greater benefits being derived from devoting resources (at the margin) to the production of CT scans, rather than to the production of, say, the services of coronary care units?

Thus the problems of putting into effect the concepts of equity are largely attributable to the problems of first defining, and then identifying, the need for a given service or procedure, as well as to securing a working definition of equity. In practice, the process is further complicated by questions such as: Is full capacity utilization indicative of true need? What proportion of utilization is generated by the supply side? How much of the utilization is limited to the purposes for which the technology was originally intended? How much additional utilization is generated by expanding the range of indications for which the technology is employed? The following section addresses these questions.

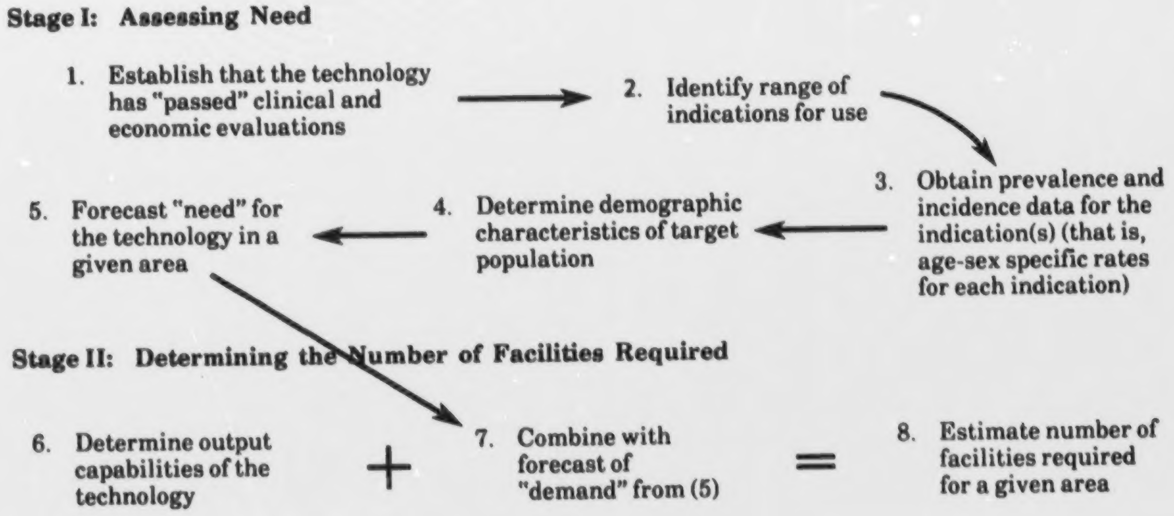
Methodology and Data Requirements for Identifying an Optimum, Using Equity as an Objective. The methodology proposed for determining the optimal number of facilities to be provided is conducted in two stages. Figure 8.1 summarizes the steps in each stage.

The first stage, the assessment of need, is often outside the purview of facilities planners. It is important, however, that planners are aware of the role of each of these steps and, moreover, are able to assess the quality of the studies used in the assessment. The importance of careful scrutiny of the data at this stage cannot be emphasized enough; unreliable and inaccurate data incorporated here will inevitably be translated into imprecise planning forecasts.

The first step in assessing need is to establish that the technology has undergone, and "passed," clinical and economic evaluations. The methodology for this step has been described in Chapters 5, 6, and 7. Planners must be aware of the integral role that this step plays in the planning process; without careful completion of clinical and economic evaluations, planners have

FIGURE 8.1

Determining the Optimal Number of Facilities


Stage I: Assessing Need
1. Establish that the technology has "passed" clinical and economic evaluations
2. Identify range of indications for use
3. Obtain prevalence and incidence data for the indication(s) (that is, age-sex specific rates for each indication)
4. Determine demographic characteristics of target population
5. Forecast "need" for the technology in a given area
Stage II: Determining the Number of Facilities Required
6. Determine output capabilities of the technology
+
7. Combine with forecast of "demand" from (5)
=
8. Estimate number of facilities required for a given area

no way of ensuring that they are providing for the optimal diffusion of a *beneficial* technology.

The second step in assessing need is to identify the range of indications for which the technology has been proven effective. Although this step may appear intuitively obvious to some, accurate and precise identification of the indications for use of the technology *at the outset* of its diffusion is crucial in ensuring that utilization is restricted to intended purposes; only then will the opportunities for inappropriate utilization be minimized.

Suppose, for example, that health care planners were assigned the task of determining the optimal number of coronary care units for a given area, assuming that clinical and economic evaluations have already been conducted. Their first step would be to verify that the admission criteria for the unit had been clearly delineated. Perhaps only patients with suspected myocardial infarction and other serious acute heart diseases would be eligible for admission; all other low-risk cardiac patients would be treated in regular inpatient facilities. Having established the indications for use, the planners are then better able to assess the amount of inappropriate utilization occurring after the unit is operative. If, for example, 40 per cent of the CCU beds were continuously occupied by low-risk cardiac patients or non-cardiac patients, then it is probable that the original estimate of the need for CCUs was high and should consequently be revised. (Routine audits and/or peer reviews are examples of ways in which utilization patterns can be monitored.) Thus, establishing definitive indications for use helps enable planners and administrators to determine the extent to which full capacity utilization of a health care technology reflects epidemiologically determined need.

Step 3 in assessing need is to obtain incidence and prevalence data for the indications identified in the first step. The finer the breakdown of indications by demographic characteristics (age, sex, etc.), the more precise the final forecast of need can be.

The next step in the process is to identify the demographic characteristics of the target population or catchment area. Since this information will eventually be combined with incidence and prevalence data from step 3, it is preferable to obtain a demographic breakdown that parallels the breakdown in incidence and prevalence rates. Demographic data often have to be adjusted (assuming that they are not already presented in the desired form) so that the total population in a given area is disaggregated in accordance with the age-sex divisions used for the incidence information.

The final step in assessing need is to integrate the information from steps 3 and 4 to derive an area-specific forecast of need for the technology in question. Of course, the estimate will be only as reliable as the data used to derive it; thus any uncertainties or imprecision in the incidence and prevalence information or demographic data should be taken into consideration at this point. This is best done by conducting a sensitivity analysis (as described in Chapter 7) around the values of the uncertain parameters. If the specific areas of imprecision in the data are not easily identifiable but rather are suspected to exist to some extent across all data, then the preferred option might be to identify a *range* of need, perhaps using confidence intervals.

Stage II then combines estimates from Stage I with additional data to determine the number of facilities required in a given area. The first step in this stage is to assess the output capabilities of the technology. Where a single technology, such as CT scanners, is involved in the production of a single output, such as scans, the productive capacity of the technology is relatively easy to determine. The planner merely has to find out how many procedures (that is, the quantity of output) are feasible in a given period and multiply this figure by the number of periods for which the technology will be in use. If, for example, a scan can be completed in one hour, and the skilled personnel required to operate the scanner are available seven hours a day, five days a week, then a hospital with one scanner can conduct at most thirty-five scans a week.

Calculations become somewhat more complicated, however, when several inputs can be combined in various proportions to produce one or more outputs.[3] In this case, information derived from a production function is necessary to identify output capabilities.[4] A production function is a technical relationship between the quantity of output of a good or service and the various quantities of inputs required to produce it. The production function, however, says nothing about the clinical value of the production process or its inputs; this is assumed to have been demonstrated in step 1.

Although the various techniques available for estimating production functions are too complex to include in the present discussion, several approaches have been used to estimate production relationships in the health care field. Among the most successful is the linear program production model, an example of which is the model developed by Hosios, Laszlo, and Levine (1978). In their study, the authors designed a production model of a diagnostic radiology department to aid in the decisions on allocation of radiographic equipment by the Quebec Ministry of Social Affairs.

Accurate assessment of the output potential of a technology is fundamental to deriving a reliable estimate of the optimal number required. The consequences of employing inaccurate information at this step are far-reaching and may even be irreversible; overestimation of a technology's capabilities will lead to excess demand and insufficient capacity, while underestimation will lead, at least initially, to insufficient utilization and excess capacity (with the consequent potential for demand generation and inappropriate utilization).

Once the productive capacity of the technology in question has been established, the next step is to incorporate this information with estimates of the corresponding demand, as derived in step 5. The reader might notice that the concept labeled need in step 5 has now become "demand." This transition implicitly assumes that all need will be realized as demand, and conversely, all demand will reflect only true need. Of course, the opportunities for both consumers and providers to distort demand beyond indications of need are numerous. The implications of these distortions are not as serious, however, *if and only if* need has been accurately measured in step 5. The reasons are as follows.

Suppose that need has been correctly evaluated, but actual utilization exceeds the estimate of need. If the planner or policy maker is prepared to allocate health care technologies based on evaluation of need, and limit their diffusion accordingly, then excess demand will likely be dealt with in triage fashion, by clinically ranking the individual merits of each indication.[5] Inappropriate demand will remain unmet, while those in need of the technology will have access to it (that is, there will be a sufficient *number* of facilities to accommodate need). Similarly, if actual demand is less than the estimate of need, then providers have the latitude to encourage additional demand in accordance with need, perhaps by disseminating information on the potential benefits of the new technology.

If, however, need has not been accurately assessed at the outset, then the implications of discrepancies between need and demand become much more severe. An underestimate of need will lead to excess demand, which will be dealt with as above, in triage fashion. The difference is that the demand that remains unmet is legitimate demand; the people denied access to the technology potentially could have benefited from it.

Just as severe in its implications is an overestimate of need. In this case the technology is supplied beyond the clinically indicated point. Two reactions are then possible. First, if actual demand reflects true need and not an incorrect estimate of need, then excess capacity will result. Supply of the technology will eventually have to be reduced – a costly and difficult process. The more likely reaction, however, is that demand will increase to the level of provision of the technology because of pressures from patients, physicians, or a combination of the two. Thus if the need is overestimated, and actual demand nonetheless equals the overestimate, then unnecessary utilization will result. That is, the technology will be provided to (and used by) people who derive little or no benefit from it; the opportunity cost of such inappropriate utilization is obviously high.

Finally, combining the output capabilities of the technology with a forecast of the demand for it yields an estimate of the optimal number required for a given area. This number is derived, however, using only need as a criterion. In reality, the provision of health care services cannot proceed unconstrained, regardless of the perceived (or actual) need for them. The fiscal limitations inherent in all health care systems, both public and private, require that priorities be assessed and services supplied accordingly. Equity considerations are one way of establishing priorities; a second, and *complementary*, method of ranking priorities is to use efficiency as an objective. The next section examines the role of efficiency in determining the optimal number of facilities.

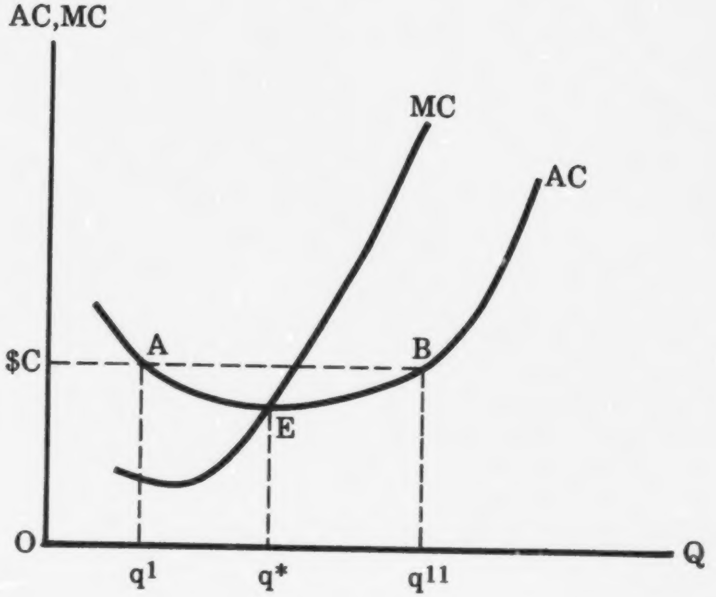
Determination of the Optimal Number of Facilities Using Efficiency as an Objective

Efficiency Defined. The concept of efficiency discussed in this chapter refers to an economic interpretation of efficiency commonly known as "technical efficiency."[6] Technical efficiency is attained in the long run when inputs are combined in a way that minimizes the cost of producing output; as such, technical efficiency can only occur when the average cost of production is at a minimum. Marginal cost, the change in total cost that results when output is varied by one unit, will equal average cost when average cost is at a minimum. Thus the optimal number of facilities, under this concept of efficiency, is the number that corresponds to the minimum cost per unit of output (as depicted by q^* in Figure 8.2).

Methodology and Data Requirements for Identifying an Optimum, Using Efficiency as an Objective. It should be evident from the definition of efficiency that cost data are necessary for determining the optimal number of facilities. But costs themselves are important only insofar as they are required for

FIGURE 8.2

Cost Curves



AC = Average Cost
MC = Marginal Cost
Q = Output

estimating the shape of the curves, knowledge of which is imperative for determining whether economies of scale exist for a given technology.

Economies of scale are said to exist when expansion of the scale of productive capacity causes total production costs to increase less than proportionately with output. Therefore, if economies of scale exist, the long-run average costs of production continuously fall until they are at a minimum point, which corresponds to the optimal output. Expansion of output past this point causes average costs to rise, which indicates the presence of *dis*economies of scale.

Economies of scale are generally classified into two groups: internal economies and external economies. Internal economies of scale are associated with the expansion of a single facility; the discussion of the sources for these economies is therefore deferred to the section on optimal size. External economies, however, are associated with the expansion in scale of the entire industry. For example, if a *single* coronary care unit became more efficient by perhaps doubling the number of beds it had, it would be benefiting from internal economies of scale (that is, the improvements in efficiency are unique to that unit). If, however, the efficiency of all CCUs was increased because, for example, the cost of monitoring units fell, external economies of scale would be responsible.

It is important for health care planners to know the shape of the average cost curve for the technology in question in order to determine whether economies (or diseconomies) of scale exist and can therefore be exploited (or avoided). The long-run average cost curve, though, is composed of a series of costs, each one corresponding to a specific output. Unfortunately, most aggregate cost data provide just one estimate of average cost, that which pertains to the output level at which the service is (or will be) provided. The problems of interpreting a one-point estimate are best illustrated by referring to Figure 8.2. Suppose a planner was provided with information that the average cost of a technology proposed for expansion was $C. Without knowing the shape of the average cost curve, the planner has no way of knowing whether $C corresponds to point A or point B (or any other point) on the curve. If at point A, expansion would result in economies of scale being realized and average cost falling toward a minimum. If, however, $C corresponded to point B, then expansion would cause increasing average cost because of diseconomies of scale, as well as a move away from the optimum. Obviously, more information is needed to reach a decision. The planner has (at least) three options for augmenting the stock of information.

The first option is the most ambitious, and probably the most unrealistic, given the time constraints with which planners often have to contend. It involves estimating cost functions from raw data, an arduous task for even those with a surfeit of time. Some work has been done in the area of hospital cost curve estimation that tends to support the hypothesis of a U-shaped curve (Hurst, 1979), but it is generally felt that the methodologies used for estimating industry cost curves are not always sufficiently sensitive to detect economies of scale.

The second option available to planners is to do a "quick and dirty" adaptation of the first option. This option avoids estimating the entire average cost curve and instead concentrates on examining the portion in the immediate vicinity of the original point estimate. The rationale for this is that no benefit derives from knowing the entire shape of the curve when only the *slope* of the curve at the point estimate of average cost is relevant for decision making. In Figure 8.2, suppose that $C, for example, was known to be the average cost of providing a certain technology. The average costs of producing slightly fewer units of the technology, as well as several more units, should then be calculated. If the average cost of producing less is higher than the original point estimates, and the average cost of producing more is lower, the slope of the curve (in that area) is negative, and $C would correspond to a point such as A. If the opposite were true, point B would represent the actual situation. The advantage of this method is that only two additional average cost estimates need be derived (as opposed to an entire schedule); the disadvantage is that although the information obtained is sufficient to aid in making decisions on whether to expand or contract output, nothing can be said about the magnitude of the change required. Planners may know, for example, that present production corresponds to a point such as A, but they have no way of knowing how far point A is away from the optimal point, E.

The third option available is to obtain additional information, not on average costs, but on marginal costs.[7] Recall from Figure 8.2 that marginal cost is below average cost when producing at less than the optimum, and above average cost when producing beyond the optimum. Thus a comparison of average and marginal costs will provide insight into whether production is occurring above or below the technically efficient point (although, once again, nothing can be said about the absolute distance from that point).[8] The study referred to in Chapter 7, for example, evaluated the cost-effectiveness of six sequential tests for the detection of asymptomatic colonic cancer. The *average* cost per case detected in six tests was $2,451, but the *marginal* cost per case detected by the sixth series was more than $47 million. This is an extreme example of marginal cost exceeding average cost, and a reduction in output is consequently indicated.

Two practical problems arise when attempting to estimate cost functions for health care technologies. The first is the dynamic nature of technological innovation; the costs estimated at the outset of the diffusion of the technology may have changed considerably by the time the technology is in widespread use. The second problem is that frequently very little cost data are available for a new technology. Moreover, the data that are available are often derived assuming that the technology will operate under ideal conditions. That is, there is often insufficient allowance for costs incurred in the set-up, day-to-day operation, and maintenance of the technology.

There are no simple solutions to these problems. Suggestions can be made, however, for avoiding them in the future. With respect to the first problem, it is essential that economic (as well as clinical!) evaluations be ongoing. Planners should not rely solely on cost estimates, which, because they are static, are outdated. The more frequent and regular the economic evaluations, the more reliable the estimates will be.

The second problem can be minimized if more management trials, rather than explanatory trials, are used to derive cost data. The costs thus obtained will more accurately reflect the actual costs incurred in practical implementation.

Incorporation of Equity and Efficiency into the Determination of the Optimal Number of Facilities

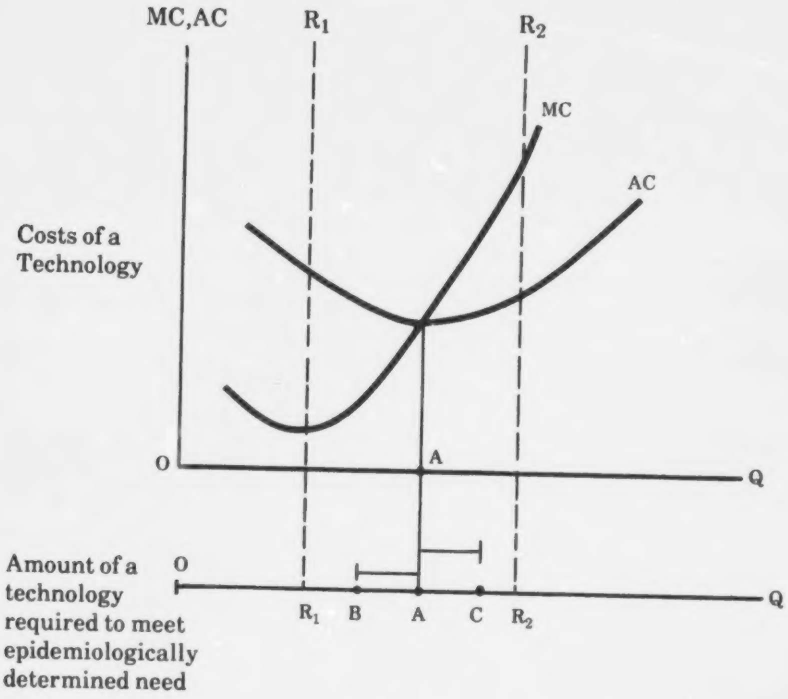
The previous sections have demonstrated how an optimum number of facilities could be determined using either equity or technical efficiency as an objective. But it was also emphasized that the methods employed are complementary and therefore should be *jointly* used to determine an optimum.

It would certainly be helpful if both equity and efficiency considerations yielded the same estimate of the optimal amount of a technology to be provided. The two objectives will probably, however, yield conflicting estimates. The various possibilities are shown in Figure 8.3, and summarized below.

The first, and most unlikely, possibility is that both equity and efficiency criteria will yield the same estimate of the optimum, say, point A. Planners then must evaluate the available resources (that is, the budget) and plan accordingly. If

FIGURE 8.3

Comparing Output Determined by Need and by Technical Efficiency



AC = Average Cost
MC = Marginal Cost
Q = Output
R = Budget Constraint

the budget is limited to an amount that will permit provision of a quantity of the technology equal only to R_1, that is the level at which it will be provided. If the budget is sufficient to purchase R_2 of the technology, rational planners will purchase a quantity of the technology equal to OA and devote the excess budget (the distance AR_2) to its next best alternative use.

The second possibility is that the estimate of the optimum derived using efficiency criteria (point A) exceeds that derived from equity criteria (point B). In this case provision of services at point A would minimize the cost per unit of service provided, although the services would be consumed by people who have no need for them—obviously not a desirable situation. Therefore provision of services would logically be limited to point B, where need is exactly satisfied, even though average costs are higher than they would be if output were expanded. Again, if resources are sufficient to provide the technology only up to R_1, OR_1 will be the quantity supplied. If the budget permits provision up to R_2, OB will be provided, and the surplus of BR_2 directed to the next highest priority. In this case, the need-determined optimum takes priority over the cost-minimizing optimum.

The third and perhaps most interesting possibility occurs when equity-derived estimates exceed the optimum calculated from an efficiency standpoint. No unequivocal solution exists to this dilemma, and it remains a source of controversy for two opposing schools of thought. Suppose that the efficient optimum was point A, but the equitable provision of services occurred at point C. Perverse as it may seem, the decision for planners is easier when the budget is constrained to the point where they are able to provide only R_1 amount of the technology. When the budget is sufficient to provide the technology in the amount of R_2, a conflict emerges.

Technical efficiency criteria suggest that services in the amount OA should be provided, allowing the surplus AR_2 to be channelled to its next best alternative use. Provision beyond point A would represent a technically inefficient use of resources in that cost per unit is higher than it would be at a lower level of provision. Equity criteria, however, suggest that the technology should be provided in the amount of OC, with the surplus of CR_2 being redirected to the next highest priority. The rationale for provision up to OC is that as long as *some* benefit can be derived by providing more of the service (that is, as long as unmet need exists), supply should be expanded to the level defined by epidemiologically determined need.[9]

The area on which the debate focuses is thus the distance AC. From the point of view of technical efficiency, AC is a measure of inefficient resource use if supply is expanded to OC; from a purely medical viewpoint, AC is a measure of unmet need if supply is constrained to OA. Planners are therefore faced with deciding how much, if any, expansion should occur past OA.

At this point, additional information is required to determine optimal output. Planners must consider not only the additional costs of expanding beyond OA, but also the additional, or marginal, benefits derived from the additional services provided. Careful comparison of the extra costs and extra benefits associated with provision in the area AC is required—both in their own right and compared with the competing alternative uses of the available resources. This is where the techniques of economic evaluation presented in Chapter 7 and demonstrated in Chapters 9 and 10 become essential for the planning process. Economic evaluation methods such as cost-effectiveness analysis, cost-benefit analysis, and cost-utility analysis provide planners with a systematic approach to assessing the relative merits of competing technologies, both their costs and their benefits. Priorities can then be assigned and decisions made about the optimal amount and *mix* of new technologies that should be provided.

The Optimal Size of Facilities

The performance and productivity of a facility is often a function of its size. The planning of individual facilities must therefore include size considerations, although size (or capacity) is obviously a more relevant consideration for some technologies than for others. In planning the diffusion of digital diagnostic imaging units, for example, the issue of how many units to locate in *one* hospital does not arise. When setting up a CCU, however, the bed capacity of the unit is a major consideration.

Many of the problems associated with determining optimal size occur because the capacity options for each facility are often discontinuous. Facilities are typically augmented in discrete increments, owing to the indivisible nature of some inputs. This implies that, for example, the decision to provide a CT scanner to an institution is a zero-one proposition; either the institute gets the CT scanner or it does not. But what happens if an assessment of need indicates that the hospital should be provided with 0.8 of a CT scanner or 8.3 CCU beds? The techniques used to determine optimal size were designed to circumvent this problem. They will be discussed in the next section, as will issues of equity and technical efficiency in capacity determination. Equity and technical efficiency will not be redefined here; the definitions found in the preceding sections are still applicable. Rather, this section will concentrate on the interpretation of these concepts in the context of planning for optimal size.

The Role of Equity in Determining Optimal Size

Recall that the objective of equity in determining the optimal number of facilities was to ensure equal access across all categories of need. The counterpart to that objective with respect to optimal size determination is to ensure equal *quality* of care for all users of the service. Therefore equity is a consideration only if quality of care varies with size of the facilities.

Several studies have examined the relationship between the volume of procedures conducted at a facility and the outcomes achieved. These studies are generally consistent in their findings of a positive relationship between volume and patient outcome (Flood, Scott, and Ewy, 1984a), even when the results are adjusted for case-mix differentials (Luft, Bunker, and Enthoven, 1979). It has been suggested that this positive relationship emerges because volume measures act as proxies for the experience and expertise of the personnel conducting the procedures; as the work team gains experience with a specific process, the chances of successful outcomes are increased.

The relationship between *size*, volume, and outcome is more difficult to resolve, however. Size is typically positively correlated with volume, and results have generally indicated that larger or more concentrated facilities exhibit, on average, better patient outcomes for selected procedures (McGregor and Pelletier, 1978; Finkler, 1979). Only one study (Flood, Scott, Ewy, 1984b) has suggested that greater size might be associated with poorer outcome after accounting for volume.

Attempts have also been made to determine a *minimum* size that would generate adequate turnover to guarantee maintenance of the skills of the providers of care. Minimum capacity levels have been recommended for cardiac surgery units, cardiac-catheterization laboratories, and facilities for coronary angiograms; as well, guidelines exist for many other diagnostic and therapeutic technologies. The job of health care planners is considerably facilitated when these minimum capacity recommendations exist. But quite often, although a quantifiable relationship between size and quality is known to exist, no minimum levels of provision have been established. Consider, for example, the findings of Bloom and Peterson (1974) on the relationship between death rates and the size of coronary care units, summarized below.

Quintile*	No. of Patients with Infarction	Deaths from Infarction	
		No.	%
1	1,780	241	13.5
2	964	193	20.0
3	614	145	23.6
4	287	71	24.7
5	133	33	24.8
Total	3,778	683	18.1

* Each quintile contains six hospitals.

Source: Bloom and Peterson (1974), p. 1172.
Reprinted by permission of the *New England Journal of Medicine.*

The inverse relationship between size and mortality rate is apparent;[10] all hospitals that had mortality rates of 20 per cent or greater treated fewer than 200 patients in the year of study. Presented with the above information, planners might decide at which point the trade-off between quality and size is acceptable. (Based on the above data, planners would likely choose to have all CCU units the maximum size. They must simultaneously consider the distribution of the units and the optimal number, both of which might exert constraints on the feasible size.) Quality considerations alone, however, are not sufficient to render a decision; issues of efficiency must also be included.

The Role of Efficiency in Determining Optimal Size

The objective of efficiency in planning is to determine the technically efficient quantity of output, as defined by the intersection of the marginal cost curve and average cost curve at its minimum point. The costs considered in this section, though, are not capital costs but *operating* costs. Because operating costs vary with size, this section will examine the choice of the size (or scale) of a facility that will minimize operating costs. (In economics jargon, the purpose of the section on the number of facilities was to determine the optimal *industry* size; this section, on the size of facilities, is devoted to determining optimal *plant* size.)

As with determining an optimal number of facilities, deciding upon the size of each facility requires knowledge about whether economies of scale exist and can be exploited. The economies of scale that occur when output of a single facility is expanded (or contracted, if operating at greater than the optimal size) are known as internal economies. If these economies exist over a very large range of outputs, then the optimal size of the facility will likely be large. The opposite is true for internal economies that are quickly exhausted at low outputs.

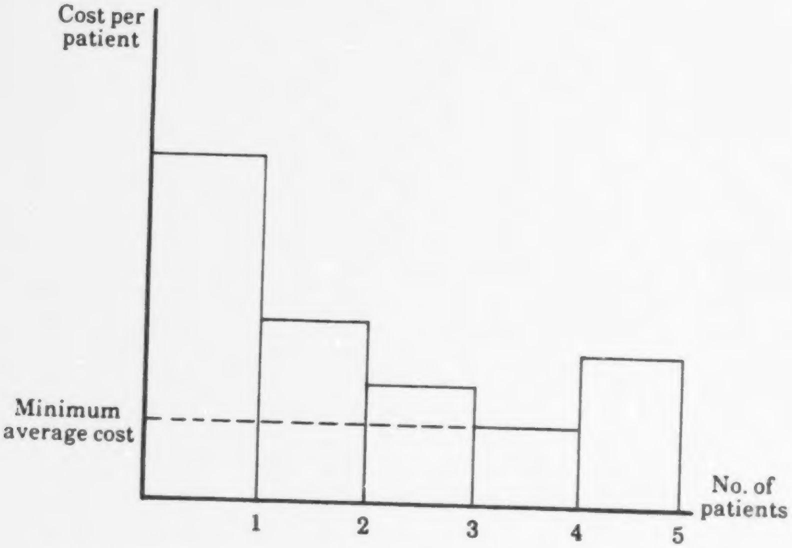
There are various reasons why internal economies, or decreasing unit costs, exist when the size of the operation is increased. First, most facilities have some amount of overhead processes that must be undertaken, regardless of the amount of output. Purchase of a defibrilator, for example, is necessary for all CCUs regardless of whether they house one patient or ten patients. But as the number of patients in the unit increases, the cost per patient for use of the defibrilator falls. In other words, as output rises, *average fixed costs* fall; and because these costs are a component of average total costs, the cost curve in Figure 8.2 is downward sloping, at least over some range.

Second, certain types of production processes may be viable only at certain rates of output, and thus as scale is increased toward that output, average costs fall. This source of internal economies arises from indivisibilities in production, which means that economies of scale would not occur if the size of the facility was capable of being augmented or decreased by very small increments. Suppose, for example, that a dialysis machine is capable of servicing four patients every week. The hospital in which it is used has, however, only one dialysis patient. In this case, the cost *per patient* of the machine is equal to the entire cost of operating the machine. If a second patient were added to the program, the average cost would fall by 50 per cent and would continue to fall if a third patient were added. Were a fourth patient to be added, the average cost would be at a minimum, because the machine would then be operating at full capacity, as shown in Figure 8.4. If a fifth patient were added, however, average cost would rise, since an additional dialysis machine would have to be purchased. Of course, these considerations of average cost would not be relevant if one-fourth of a dialysis machine could be purchased at a time; it is precisely the *indivisible* nature of the machine that causes internal economies of scale to exist.

An indivisibility especially relevant for technology diffusion is the minimum amount of a technology necessary for a teaching and research facility. Suppose, for example, that a university hospital requires a CT scanner for teaching. If there is not sufficient demand to operate the scanner at full capacity, the average cost per scan will not be at a minimum. Suppose, though, that the catchment area the hospital serves is expanded. As use

FIGURE 8.4

Cost per Patient for Dialysis Machine



of the scanner increases, the cost per scan will fall, and internal economies of scale will be realized. The teaching and research facility is therefore especially vulnerable to inefficiencies resulting from indivisibilities of technologies because, barring generation of demand for its services, it has little opportunity to exploit potential internal economies from this source.

A third source of internal economies arises when expansion in the scale of facilities permits greater specialization and division of labour. When tasks can be divided according to specialization, greater productivity can be obtained from the labour inputs, thereby causing a decrease in average costs.

In determining optimal size, therefore, it is important to assess if, and where, economies of scale exist. This exercise is considerably easier for individual facilities than for the aggregate, the industry as a whole, because institution-specific cost data are relatively easy to access.

Within the health care literature, several studies attempt to measure average costs of various technologies to determine if and where internal economies of scale exist. Finkler (1979), for example, has found that the average cost for cardiac surgery is approximately $15,000 when 100 operations are performed per year at a given facility. This figure drops to $9,000 when 400 operations are performed; thereafter, average costs do not vary significantly with output, and the average cost curve flattens out. Bartlett and colleagues (1978) also found a downward sloping average cost curve for brain scans at Brook General Hospital in London. The cost per scan fell from £73 when 600 scans per year were performed to £20 when 2,600 scans were done. Again, the curve flattened out at 2,600 scans; the average cost per scan was no different whether 2,600 or 3,000 scans were performed.

The techniques for estimating average cost functions for an individual institution require that institute-specific cost data be obtained. The purchase price of capital equipment is available from the manufacturer, and costs of operating the technology (including personnel costs, costs of supplies and equipment, overhead expenses and depreciation) are available from hospital sources (assuming of course, that the technology is housed in a hospital). A variety of techniques exist for estimating institute-specific costs, ranging from quick and dirty "guesstimates" to sophisticated step-down, overhead cost allocation methods.[11] Not surprisingly, it is usually the case that the more sophisticated and thorough the costing techniques, the more confidence the planner will be able to put in the resulting estimates.

Caution should be exercised, however, when interpreting the results of cost analyses conducted using the case-study approach. Many of the cost estimates derived with this method are dependent upon the characteristics of the institution or setting in which the analysis was conducted. As such, the results cannot necessarily be generalized to facilities in which the case study conditions do not prevail. A cost-effectiveness analysis undertaken by Finkler (1979), for example, indicated that the technically efficient production of heart surgeries requires a minimum volume of 500 procedures per centre per year. When the study was replicated in a different setting, however, efficient production was attained at 300 cases per year (Finkler, 1981). Thus, it is essential when estimating cost functions to utilize data that are specific to the facility in which the technology will be housed.

Given knowledge of the costs at various levels of output, the planner must then incorporate information on equity with that on efficiency to determine the optimal size of the facility in question.

Incorporating Equity and Efficiency Objectives into Decisions on Optimal Size

Optimal size as determined by quality considerations and optimal size as determined by cost considerations are not necessarily equal. The planner is thus faced with the task of seeking the best mix of quality and costs, given the constraints on existing resources.

The most widely used technique for incorporating considerations of both quality and capacity into facilities planning is a mathematical approach known as linear programming. Broadly defined, linear programming is "set of mathematical techniques for finding numerical solutions to a special class of problems . . . An essential part of the problem will be the existence of inequality constraints or limitations on the values of the variables which may be selected. The problem is then to choose the best possible values out of those permitted by the constraints" (Bannock, Baxter, and Rees, 1982). The linear programming approach begins by defining an objective function – a mathematical expression that contains quantitative information on the variables to be optimized (such as capacity). The second element of linear programming consists of a set of inequality constraints (such as cost) that bound the range of feasible solutions for the objective function. The mathematics of linear programming typically involves searching for a solution that minimizes the objective function subject to the existing constraints, although, depending upon the nature of the variables in question, the optimum may be determined by *maximizing* the objective function.

The constrained optimization technique for determining optimal size is receiving increasing attention from planners of health care facilities (see, for example, Greenwald,Woodward, and Berg, 1979). The distinct advantage of this approach is that it provides a systematic method for simultaneously including cost and quality information in capacity decisions. Planners might, for example, establish as their objective function the cost equation for the particular technology. (A cost function specifies the quantitative relationship between input requirements and output potential and identifies the associated costs, as determined by the cost coefficients or unit cost of each resource.) The next step is then to determine the constraints, in terms of, say, a minimally acceptable level of provision (from a quality perspective), which bound the solution set. The linear programming approach searches for an optimal solution that will minimize cost (that is, find the closest point to the minimum of the average cost curve) subject to the limitations imposed by the constraints (for example, that a minimum of 200 procedures be performed each year). Alternatively, planners might opt to maximize output, subject to a range of acceptable costs.[12] (This option would be preferable if, for example, guidelines indicated that quality was maximized when output was maximized.) As well, a specific form of linear programming, known as integer programming, ensures that the solution is restricted to non-fractional possibilities. This restriction precludes obtaining an optimal size solution, which, for practical purposes, is not feasible due to the indivisible nature of the technology (for example, 0.8 of a CT scanner).

Linear programming is, however, not without limitations. Because the constraints must always be precisely specified, it is necessary to have accurate and reliable data on *either* costs or quality considerations. If planners want to minimize cost subject to a minimally acceptable level of output, existing guidelines must dictate what that minimum output should be. Similarly, if output is to be maximized subject to minimum cost constraints, information on where (that is, at what output) average costs are minimized must be available. In other words, linear programming does not provide a method for *simultaneously* determining optimal size from a cost perspective and from a quality perspective; one of the optima must be specified a priori.

As well, it is important to remember that linear programming is merely a method of analyzing existing data. As such, it cannot create new data or enhance the quality of questionable data. Thus, as with other techniques available for optimal facilities planning, linear programming will produce estimates that are only as accurate as the data used to derive them.

Although capacity options are sometimes evaluated according to cost and quality considerations alone, it is more often the case that the location of the facility will also have an influence on the optimal size. Consequently, determining the optimal number and the optimal size of facilities are often viewed as necessary first steps undertaken prior to, or concurrently with, the determination of optimal location. It is this final issue to which we now turn.

The Optimal Location of Facilities

Determining the distribution of health care technologies completes the logical sequence of optimal facilities planning. Decisions on the spatial allocation of facilities are typically assumed in theory to occur only after the optimal number has been established; thus location planning is essentially a matter of devising the best distribution of a predetermined number of facilities, as defined by some generally accepted criteria. (However, it may be the case that the optimal number, or total provision, of a technology is determined by a prior assessment of what its optimal distribution should be.) As with determinations of optimal number and optimal size, locational decisions include considerations of both equity and efficiency.

The Role of Equity in Determining Optimal Location

An equitable distribution of facilities is one in which equal *access* to care is ensured, regardless of the geographical location of the client group. Because the number of users and frequency of their use of the technology varies inversely with the distance they must travel to seek the service, the greater the travel time or distance, the greater the barrier to access. Patients associate a certain amount of disutility with time and money expended in obtaining access to a service. One objective of facilities planning is, therefore, to determine a distribution that minimizes the aggregate disutility resulting from the spatial allocation of technologies or, alternatively, that maximizes public welfare. Moreover, because demand for the technology (on the part of both patients and physicians) is negatively related to travel distance, underutilization may result if the technology is perceived to be inaccessible.

Attainment of this objective is confounded considerably by the typically uneven distribution of potential users of the technology over an area. Consequently, consolidation of services in a populous region would disadvantage potential users in outlying areas, while dispersion of services to outlying areas would force users in densely populated core areas to travel farther than they would have to if the services were centrally located. Planning for the equitable distribution of a technology must therefore acknowledge and accommodate the spatial, temporal, and random distribution of demand for the technology.

The Role of Efficiency in Determining Optimal Location

A technically efficient distribution of facilities is one in which the services are rendered at minimum cost. Both capital and operating costs are a function of location; the more remote and isolated the location, the higher the costs of building and/or operating the facility will be. This observation has led to a call for increased centralization of facilities in the hope that the concentration of specialized services at fewer, centrally located facilities would reduce the aggregate cost of providing the service. Advocates of this argument suggest that when expensive technologies are widely dispersed, inefficiencies result from both underutilization of facilities in relatively sparsely populated areas and from duplication in the provision of infrequently needed services. As well, it has been suggested that the presence of many competing centres capable of carrying out a certain procedure may stimulate a level of utilization which is greater than that of a single centre with the combined capacities for that procedure and, more important, which is greater than that required to satisfy need. Thus the debate over centralization is further confounded by the possibility that the total number of procedures is likely to be influenced by the dispersion or concentration of facilities.

Unlike optimal number and optimal capacity decisions, capital and operating costs are not the only costs that must be included in the efficiency considerations of location planning. Patients' access costs are also a significant component of the total cost of providing the technology. These costs include items such as travel expenses, income losses from lost production time, child care and other expenses.[13] These costs are fundamentally different from capital and operating costs, not only because of their nature and who bears them, but more importantly because, unlike other costs, they vary *inversely* with the dispersion of the facilities; that is, the more facilities available, the less the patient has to travel to reach the closest one. Thus, planners must include in their location decisions an evaluation (that is, enumeration and comparison) of the marginal costs to patients versus the marginal savings from centralization.

Incorporating Equity and Efficiency Objectives Into Optimal Location

Optimal location decisions are considerably facilitated by the multitude of methodologic techniques available for addressing the problem. Most of the methods are highly technical, thereby precluding a detailed exposition of them in this chapter. Instead, a descriptive summary of some of the most widely used methods is provided in an effort to identify the basic attributes and distinguishing characteristics of each.

By far the most common approach to location determination is one in which only equity considerations are included; specifically, most techniques are designed to provide a distributional solution that minimizes aggregate travel time or distance (that is, the sum of *all* patients' travel time or distance to *all* facilities). As with all optimal location techniques, this procedure requires knowledge of the population density, as well as data on the distribution of need or anticipated utilization for the area (be it a county, province or country) in which the technology is to be diffused. Once this information is mapped, isochronic lines – lines of equal travel time (or distance) – are drawn around all possible candidates for technology locations (a large number of possibilities usually requires computer mapping). The number of people falling within each isochrone is then calculated and used to derive the total travel time (or distance) associated with each potential facility location. The optimal solution is the one that yields the minimum total travel time for the population.

Numerous variations of this approach appear in the literature and have recently been reviewed by Revelle and colleagues (1977) in the context of their applicability to emergency medical services. The location decision problem was first modelled as a "p-median problem," which defines the objective of facilities planning as an attempt to "locate p-facilities on a network of demands so that average travel time of users is a minimum" (Revelle et al., 1977). Mathematically optimal solutions to the p-median problem typically are obtained using linear programming. This technique was then expanded to "chance-constrained programming," an approach that extended the p-median concept to the situation in which demand for service in a particular area arises randomly with some mean and variance. The p-median problem was subsequently modified by adding maximum time constraints to ensure that no point of demand would be left with its closest facility farther than some (pre-determined) maximum time away. From this modification arose the "location set covering model," which seeks to find the *minimum* number of facilities required, given the explicit time constraints governing choice of location. This model gradually evolved into what is known as the "maximal covering model," which is the latest innovation in location theory. An example of this approach is provided in the technical appendix.

The maximal covering solution is an example of how considerations of travel distance and population coverage can be integrated into decisions on facility location. Practical application of this technique has resulted in a number of variants most easily distinguishable by the way they integrate expectations of demand into the model. One approach, pioneered by the Swedish Planning and Rationalization Institute (SPRI), is to base forecasts of future utilization trends on evidence of past patterns, as defined by prevailing incidence rates for the disease in question. An example of this process applied to the planning of CCUs is found in Jonsson (1981). First, population densities were established for the area of consideration, in this case, Sweden. Next, age-specific and sex-specific rates of heart disease were calculated, based on data from both Sweden and the United States. These data were then combined to predict the flow of patients to CCUs. Finally, the estimates of demand for the service were used to derive a distribution of CCUs that minimizes aggregate travel time to the units.

Although the approach has received considerable attention from SPRI, it has not been adopted with equal vigour by other planning agencies (few of which are as sophisticated in their approaches as SPRI). This may be, in part, because the approach assumes that past utilization trends are indicative of future ones or that they reflect appropriate patterns of utilization, even though they may have been the result of past unplanned activities. Furthermore, no allowance is made for the influence on utilization of differing socio-economic factors among subregions. An alternative technique, which takes into account the omissions from the SPRI approach, relies on a heuristic simulation of health facility location decisions. Utilization of a facility is simulated using estimates of the quantitative relationships among demand and population density, average income, age, and other pertinent demographic factors (Calvo and Marks, 1973). This approach has the advantage of allowing anticipated changes in the demographic features of the population to be incorporated into forecasts of future utilization.

The role of socio-economic factors on utilization has been explored and quantified in a spatial allocation model for the planning of regional health services by Abernathy and Hershey (1972). In their model, the authors determine an index for measuring the sensitivity of utilization to distance from a health care facility for various social strata. Their model recognizes and incorporates evidence that distance provides a greater barrier to access for lower socio-economic strata.

All of the above models share one common characteristic: measures of access constitute the sole criterion for location decisions. The last model reviewed in this chapter goes one step farther and incorporates both equity and efficiency criteria into the optimal solution. A sophisticated form of linear program, known as multi-objective optimization, is used to search for a distribution pattern that fulfills a range of objectives. Calvo and Marks (1973) have proposed a theoretic multi-attribute objective function that simultaneously attempts to minimize distance or travel time, minimize user costs, maximize demand, and maximize utility (where utility is defined as a function of the facility's degree of utilization). An even more ambitious multi-objective optimization model has been developed by Bloom and Peterson (1974) and successfully applied to distribution of CCUs in Massachusetts. The authors identify six criteria that should apply to the optimal solution.

1. All patients should have ready access to a unit; 30 minutes travel time is the permissible maximum.
2. Every unit should be directed by a qualified cardiologist with appropriate organizational and administrative skills.
3. Units should be strategically located and of sufficient size so that every patient would have at least a 95 per cent chance of finding an available bed in the nearest unit.
4. Units should have at least eight beds so that economies of scale can be realized.
5. Each major university teaching hospital in a given area must have a unit to meet teaching, research, and patient-care responsibilities.
6. At least one unit should exist in each suburban sector of a metropolitan area.

Unfortunately, when the authors attempted to distribute CCUs according to the above criteria, they found that strict adherence to the constraints meant that no solution existed. Constraints 1 and 4 were therefore relaxed in order to find an optimum solution. Thus, in attempting to utilize this methodology, planners must be prepared to assign priorities to their constraints because it is unlikely that all objectives can be achieved concurrently and to the same degree. As well, users of these techniques might consider conducting sensitivity analyses to assess the extent to which relaxing some of the constraints alters the conlcusions about the optimal distribution of the technology.

The choice of technique for determining an optimal distribution of facilities will obviously depend upon the nature of the optimization problem, as well as the quantity and quality of data available for inclusion in the analysis. Facilities planners should be aware that these techniques do exist, have been used (though are generally underutilized), and can contribute to rational planning decisions regarding the diffusion of medical technologies.

Conclusion

This chapter has discussed many of the issues and concerns that confront health care planners when making decisions about the optimal provision and utilization of new health care technologies. It is not intended to be a "how-to" manual for planning; the technical nature of the methods introduced render it impossible to reproduce the step-by-step details of each method in the space available here. Readers are therefore referred to the references for an elaboration of the various techniques available. Rather, the chapter is intended to present the planner with a systematic framework for organizing the many considerations that should be addressed in planning for the diffusion of a new technology.[14]

Finally, it should be emphasized that these approaches serve merely as an aid to decision making, not as a replacement for it. The ultimate decisions regarding priorities and trade-offs cannot be resolved solely with an appeal to these techniques. As Abernathy and Hershey (1972) stress:

> In the final analysis, the decision maker must decide how to weigh different objectives. No one criterion is best. The ultimate decision as to the number, capacity, and location of facilities must be based upon a broad range of considerations. However, formal analysis can be extremely useful in reducing the number of variables that must be considered subjectively.

Technical Appendix

The maximal covering location problem is to "allocate facilities to positions on the demand network so that the maximum population will find service within a stated time or distance standard" (Revelle et al., 1977). Figure 8.A1 provides an example of the type of solution that the maximal covering model might yield. Specifically, it illustrates the proportion of population covered by a set of optimally located facilities within a distance standard. The facilities (circled dots) are located so that all demand points (solid dots) and 100 per cent of the population are within 15 unit limits (solid lines) from some facility. (Units may be measured in time or distance.) Dashed lines are 10 unit limits; they enclose 55 per cent of the population.

Notes

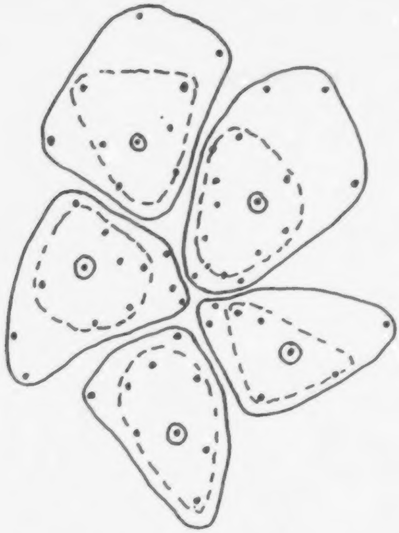
1. To date, discussions about optimal facilities planning in health care typically have concentrated on how to disperse facilities such that travel distances for the users of the facilities are minimized, subject to a variety of cost constraints. As discussed below, however, travel time and distance is but one consideration when planning for the provision of health care facilities. The standard definition of optimal facilities planning is therefore expanded for the purposes of this book.

2. See, for example, Boulding (1960), Bradshaw (1972), Culyer (1976), and Williams (1978) for discussion of the spectrum of definitions of need.

3. Of course, the example of the CT scan really belongs in this category, since labour is a necessary input in producing the output scans. That is, at least two inputs are requisite for output: labour in the form of skilled personnel and capital in the form of the scanner. In fact, labour is required for the operation of most technologies. But some labour inputs cannot be varied, as for the scanner where a minimum amount of a trained technician's time is necessary for operation. Other labour inputs, however, can be combined in varying amounts to produce certain outputs (for example, well-baby visits can be conducted by physicians, nurse practitioners, or a combination of the two). The latter case is what is meant in the text by "two or more inputs."

FIGURE 8.A1

A Maximal Covering Solution Location Problem



Source: Revelle et al. (1977), Figure 3, p. 139. Reprinted with the permission of the Hospital Research and Educational Trust from *Health Services Research*.

4. In general the production function can be written in the form $Q = f(K,L)$, where K represents all capital input used in the production process, and L represents the labour input. Technical change may affect the nature or quality of the capital input, or the labour input, or both. It may also alter the whole production function, allowing more output to be produced from the same amount of inputs. The characteristics of a technology and its production potential are thus summarized in the production function relationship and can be used to estimate the cost of producing various amounts of output.

5. An example of the triage approach to rationing scarce resources is found in Singer and colleagues (1983). The authors analyzed the effect of a cutback in beds in an intensive care unit and concluded that "moderate resource limitations can encourage more efficient use of intensive care."

6. The concept of efficiency utilized in this chapter will always refer solely to technical efficiency; the notion of "allocative efficiency" often cited in the economics literature is not included in the definition. Consequently, the term "efficiency" does not exactly correspond here with the "efficiency" concepts discussed in Chapters 7, 9, 10, and 11. The latter usage concentrates on comparing the *relative* efficiency of two competing alternatives; in this chapter, however, efficiency deals with examining the production process of a *single* technology, and as such is not concerned directly with allocative issues.

7. Although no distinction has been made in the text between short-run and long-run costs, it is assumed that all cost curves to be estimated are long-run curves. Notwithstanding the possibility that this assumption may be violated in reality, it is adopted here because it has the attractive feature of allowing all productive inputs to be varied. (It also reflects the real world situation with respect to introduction decisions about technology, and perhaps even with respect to expansions and contractions, because the lifetime of some machines or procedures may be quite short. Therefore, the term "long-run", although primarily an analytic concept, can be interpreted as the "not-too-distant" future, at least in planning.) This flexibility in factor proportions is especially important for the

measurement of marginal costs; if measured in the short run, marginal costs depend only on variable costs (since fixed costs are, by definition, fixed in the short run). Thus if marginal costs are to reflect the rate of change in *total* costs, only the long run is relevant.

8. This comparison will be valid only if the average cost curve has the traditional U-shape and the marginal cost curve is upward sloping throughout the relevant range. Situations where average cost is everywhere equal to marginal cost or where marginal cost always remains below average cost will obviously not be candidates for this comparison. Unfortunately, the only way to determine if the above situations exist is to estimate the entire curves. Fortunately, these situations are uncommon enough not to warrant serious concern here.

9. Because point C is based on an epidemiologic assessment of need, the planner is assured that the marginal benefit of providing the service is still non-negative at C. (If perfect information existed on the effectiveness of the technology, the indications for its use, its output capabilities, the prevalence and incidence of disease, and the demographic characteristics of the population, need would be perfectly identified. Consequently, the marginal benefits of additional provision of the technology would, by definition, equal zero at point C.) This would not necessarily follow if C were defined on the basis of demand; in that situation, the marginal benefits at C could easily be negative, and expansion to C would be unlikely to represent a socially optimal allocation of resources.

10. As always, users of risk-benefit (or any other evaluative studies) should view the results with a critical eye. The authors of the above study acknowledge, for example, that some of the differences in death rates may be due to patient selection. For a complete exposition of how to appraise critically a clinical study, see Department of Clinical Epidemiology and Biostatistics (1981).

11. See Boyle and colleagues (1982) for a detailed example of a sophisticated approach to cost allocation.

12. In this case, the constraint specifies an upper limit rather than a minimal level of provision. Greenwald, Woodward,

and Berg (1979) suggest that use of a capacity constraint approximating upper operating limits, as opposed to acceptable minimum levels of operation, is preferable because it simulates capital expenditures more realistically.

13. The psychic costs of patient inconvenience, family disruptions, etc., should theoretically also be included. Because they are extremely difficult to quantify, however, they have been omitted here.

14. Although the basic principles discussed here are relevant in a variety of settings, the specific techniques used will depend on the planning questions being asked and the setting within which the planning is being conducted. In the health care sector, planning is typically done at a variety of levels including the hospital or other health care providing organization, the district health council, the provincial ministry of health, and the federal health ministry.

References

Abernathy, W.J., and Hershey, J.C. 1972. "A Spatial-Allocation Model for Regional Health Services Planning." *Operations Research* 20: 629-642.

Bannock, G.; Baxter, R.; Rees, R. 1982. *The Penguin Dictionary of Economics*, 3rd ed. Middlesex, England: Penguin Books.

Bartlett, J.R.; Neil-Dwyer, G.; Banham, J.M.M.; Cruikshank, D.G. 1978. "Evaluating Cost-Effectiveness of Diagnostic Equipment: The Brain Scanner Case." *British Medical Journal* 2: 815-820.

Bloom, B.S.; and Peterson, O.L. 1974. "Patient Needs and Medical-Care Planning: The Coronary Care Unit as a Model." *New England Journal of Medicine* 299: 1171-1177.

Boulding, K.F. 1960. "The Concept of Need for Health Services." *Milbank Memorial Fund Quarterly* 44: 202-223.

Boyle, M.H.; Torrance, G.W.; Horwood, S.P.; Sinclair, J.C. 1982. *A Cost Analysis of Providing Neonatal Intensive Care (NIC) to 500-1499 Gram Birth Weight Infants.* QSEP Research

Report No. 51, Program for Quantitative Studies in Economics and Population. Hamilton, Ontario: McMaster University.

Bradshaw, J. 1972. "A Taxonomy of Social Need." In *Problems and Progress in Medical Care: Essays on Current Research,* 7th series, ed. G. McLachlan. London: Oxford University Press.

Culyer, A.J. 1976. *Need and the National Health Service.* London: Martin Robertson.

Calvo, A.B., and Marks, D.H. 1973. "Location of Health Care Facilities: An Analytic Approach." *Socio-Economic Planning Sciences* 7: 407-422.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1981. "How to Read a Clinical Journal." *Canadian Medical Association Journal* 124: 703-710, 869-872, 1156-1162.

Finkler, S.A. 1979. "Cost Effectiveness of Regionalization: The Heart Surgery Example." *Inquiry* 16: 264-270.

Finkler, S.A. 1981. "Cost Effectiveness of Regionalization – Further Results for Heart Surgery."*Health Services Research* 16(3): 325-333.

Flood, A.B.; Scott, W.R.; Ewy, W. 1984a. "Does Practice Make Perfect? Part I: The Relation between Hospital Volume and Outcomes for Selected Diagnostic Categories." *Medical Care* 22(2): 98-114.

Flood, A.B.; Scott, W.R.; Ewy, W. 1984b. "Does Practice Make Perfect? Part II: The Relation between Volume and Outcomes and Other Hospital Characteristics." *Medical Care* 22(2): 115-125.

Greenwald, H.P.; Woodward, J.M.; Berg, D.H. 1979. "Transportation or CT Scanners: A Theory and Method of Health Resources Allocation." *Health Services Research* 14(3): 207-219.

Hosios, A.J.; Laszlo, C.A.; Levine, M.D. 1978. "A Model for Support Radiographic Equipment Allocation Decisions." *Journal of the Operational Research Society* 29(3): 205-214.

Hurst, J. 1979. "Planning and Hospital Costs." In *Economics and Health Planning*, ed. K. Lee. London: Croom Helm.

Jonsson, E. 1981. *Studies in Health Economics*. Stockholm: The Economic Research Institute, Stockholm School of Economics.

Luft, H.S.; Bunker, J.P.; Enthoven, A.C. 1979. "Should Operations be Regionalized? The Empirical Relation between Surgical Volume and Mortality." *New England Journal of Medicine* 301(25): 1364-1369.

McGregor, M., and Pelletier, G. 1978. "Planning of Specialized Heart Facilities: Size vs. Cost-Effectiveness in Heart Surgery." *New England Journal of Medicine* 299: 179-181.

Mooney, G.H. 1983. "Equity in Health Care: Confronting the Confusion." *Effective Health Care* 1(4): 179-184.

Nelson, C. 1976. "Linking Health Needs and Facility Location: The Health Care Planner-Manager's Challenge." *Health Care Management Review* Spring: 29-38.

Platt, R.D. 1971. "Utilization of Facilities for Heart Surgery." *New England Journal of Medicine* 284(24): 1386-1387.

Revelle, C.; Bigman, D.; Schilling, D.; Cohon, I.; Church, R. 1977. "Facility Location: A Review of Context-Free and EMS Models." *Health Services Research* 12(2): 129-146.

Shapiro, S.H., and Wyman, S.M. 1976. "Cat Fever." *New England Journal of Medicine* 294(17): 954-956.

Singer, D.E.; Carr, P.L.; Mulley, A.G.; Thibault, G.E. 1983. "Rationing Intensive Care – Physician Responses to a Resource Shortage." *New England Journal of Medicine* ; 309(19): 1155-1160.

Williams, A. 1978. "Need – An Economic Exegesis." In *Economic Aspects of Health Services*, ed. A.J. Culyer and K.G. Wright, pp. 32-45. London: Martin Robertson.

Chapter 9:

Clinical and Economic Evaluation of Diagnostic Technologies: The Cases of Impedance Plethysmography and Nuclear Magnetic Resonance

Gordon Guyatt & Michael Drummond

Impedance Plethysmography: Clinical Assessment

The development of blood clots in the veins of the legs – deep venous thrombosis – is a common medical problem. Deep venous thrombosis (DVT) is of concern not only because of the local symptoms, such as pain and swelling, but because the clots that develop in the legs can break off and clog major blood vessels in the lung. This event, referred to as pulmonary embolism (PE), is often serious and can be fatal. The accurate diagnosis of DVT is especially important because the therapy, anticoagulation, is effective in treating the local problem and preventing PE, but potentially toxic in precipitating minor and major bleeding.

Until the past decade, the diagnosis of DVT was made on clinical grounds: when physicians observed pain, redness, and swelling in a lower extremity, they treated with the anticoagulants herapin (initially) and coumadin (often for periods of three months or more). It was discovered, however, that the clinical diagnosis of DVT is both insensitive (a limb can look normal and still have DVT) and nonspecific (in about 50 per cent of red, hot, tender, swollen legs, the cause will not be DVT), when assessed by the gold standard, venography.

Venography is performed by inserting a needle into a vein in the foot and injecting some dye that allows visualization of the venous system of the leg. Venography is painful, involves exposure to radiation, can be associated with reactions to the dye, and is not available in some smaller hospitals. It is also moderately expensive if performed as an inpatient procedure. Therefore, the discovery that the clinical diagnosis of DVT is inaccurate led to the search for a less invasive way of making the diagnosis. One such method is scanning with ^{125}iodine-labelled fibrinogen (or, simply, ^{125}I-fibrinogen), in which radioactive fibrinogen, concentrations of which are detectable by scanning with a radiation-sensitive device, is injected intravenously and taken up by actively forming clots. Another non-invasive approach to diagnosis of DVT, impedance plethysmography (IPG), will be discussed here in detail.

The electrical resistance (impedance) of the leg varies with the extent of filling of the venous system and can be measured by placing a pair of electrodes on the calf. Successive inflations and deflations of a pneumatic cuff (as is used for blood pressure measurements) placed around the midthigh leads, in normals, to rapid changes in blood volume (and thus in impedance) in the leg. In patients with DVT in the large veins of the thigh (proximal DVT), these rapid changes in blood volume do not occur, and it was hypothesized that patients with DVT might be distinguished from those who do not have DVT on this basis.

Initial work with IPG focused on establishing the best method of conducting the test. The accuracy of IPG varied in initial studies (Wheeler et al., 1974; Wheeler, 1973; Johnston, Kakkar, and Stimbler, 1974; Hull et al., 1977; Richards et al., 1976; Hull, Hirsh, and Sackett, 1979; Moser, Brach, and Dolan, 1977) but was quite consistent once a standard optimal technique (which involves serial inflation and deflation of the pneumatic cuff) was established. This discussion will focus on the work with that technique which was undertaken by a single group of investigators, led by Drs. Russell Hull and Jack Hirsh at McMaster University in Hamilton.

Accuracy

In the initial work with IPG the sensitivity and specificity of the test were determined by comparison with the gold standard, venography. The first study (Hull et al., 1976) was conducted on consecutive hospitalized patients referred because of suspicion of DVT on the basis of either symptoms and signs or positive ^{125}I-fibrinogen leg scanning, and strategies were adopted to maximize the reliability of interpretation of the gold standard (Hull et al., 1978). IPG and venography were each interpreted without knowledge of either the results of the other test or the patient's clinical condition. The method of administering IPG was described in detail in this and subsequent publications (Hull et al., 1976; Hull et al., 1978).

Each of the guidelines we have described for demonstration of accuracy were adhered to in this study. Blind reading of the test and the gold standard were undertaken to determine the sensitivity and specificity of the test in an appropriate population of patients referred because of suspected DVT (rather than, say, a group of normals and a group of patients with documented large DVT). In proximal DVT the sensitivity of the test was 93 per cent and the specificity 98 per cent. The test was not, however, nearly as good in distal DVT (clots confined to the calf), in which the sensitivity was only 17 per cent.

Hull and colleagues went on to examine the accuracy of IPG and leg scanning in outpatients referred because the attending physician suspected a first episode of DVT (Hull et al., 1981). Again, the methodological criteria we have outlined were rigorously adhered to. In this investigation, IPG was 95 per cent sensitive and 98 per cent specific for proximal DVT, but once again much less satisfactory results for distal DVT were obtained. IPG and leg scanning were administered to all patients, and taken together, however, the two tests had a sensitivity of 100 per cent and a specificity of 98 per cent. These results led the authors to suggest that the combination of leg scanning and IPG could replace venography as a technique to diagnose DVT.

Therapaeutic Effect

Because the relation between the diagnostic test result and therapy in DVT is so direct (a positive result immediately mandates ten days of intravenous heparin followed by up to three months or more of oral warfarin), separate formal studies of therapeutic effect are not necessary. Taking, for example, the result of Hull's study of outpatients referred because of the clinical suspicion of DVT, combined IPG and leg scanning would have led to all patients with DVT (according to the gold standard, venography) being treated with heparin, and to 5 per cent of those who did not have DVT being treated as well. Basing treatment on IPG alone would have led to 95 per cent of those with proximal thrombi, but only a fraction of those with distal disease, receiving anticoagulants; 2 per cent of those without DVT would have been treated.

Do these data establish that IPG and leg scanning should take the place of venography in diagnosis of DVT? In our guidelines, looking at the general case, we recommended that the effect on patient outcome be established before a new technology is adopted. This would suggest that the investigators might proceed to a randomized controlled trial (RCT) in which patients would be randomized to venography or IPG and leg scanning. Several features of present situation make this next step unnecessary, however. First, we know that heparin is effective and beneficial in DVT, not only in treatment of local symptoms, but in prolonging life through prevention of potentially fatal pulmonary emboli. Second, the results of the new technology and the gold standard are so close that the treatment people receive would be identical except in a very small fraction of those presenting with DVT (the 5 per cent without DVT who would receive anticoagulants). Third, the differential morbidity in the two groups is easy to estimate on the basis of available data (pain, radiation, inconvenience, and small risk of DVT induced by the procedure for all patients getting venograms, versus the risks of anticoagulant for the 5 per cent of those without DVT who would receive treatment if tested with leg scan and IPG). Before making a final decision whether to adopt the new technology, one would examine the economic ramifications, which we will discuss subsequently.

Based on the available information, the McMaster group felt comfortable replacing venography with IPG and leg scanning. This was not the end of the story, however, for they suspected that distal DVT was not clinically important, in that it generally resolved on its own and seldom led to clinically important sequelae, such as pulmonary embolism, without first extending proximally (in which case, the IPG would become positive). The therapeutic effect of abandoning leg scanning is clear; 2 per cent of patients with proximal DVT and about 85 per cent of those with distal DVT would not receive anticoagulants, while about 3 per cent of those without DVT would be spared treatment. The effect on patient outcome is far from clear, however. Those with DVT who are not treated may all resolve spontaneously, or a proportion may go on to have fatal PE. Obviously, if one thought that the likelihood of the latter possibility were high, it would be unethical to abandon leg scanning; because it was thought to be low, Hull and colleagues elected to conduct a controlled trial of IPG alone versus IPG with leg scanning, examining patient outcome as a result of application of the alternative diagnostic strategies.

Patient Outcome

In the latest study from Hull's group, IPG was performed in 964 consecutive patients with clinically suspected deep venous thrombosis. If the IPG was negative, eligible patients were randomized to either IPG alone or to IPG and leg scanning. In the group receiving IPG alone, the test was repeated up to day 14; in the group receiving leg scanning, patients were screened at 48 and 72 hours. Treatment was withheld in all patients who remained negative while positive tests were confirmed by venography and treated with anticoagulants. All patients were followed for 12 months.

Of the 645 eligible patients who were initially negative by IPG, 328 were randomized to IPG and leg scanning and 317 to serial IPG alone. On repeat testing during the week following presentation, 58 patients in the IPG and leg scanning group had positive findings (30 confirmed by venography) compared with 8 patients in the IPG alone group (6 confirmed by venography), a difference which is statistically significant at a 0.01 per cent level of significance. During the long-term follow-up, seven patients in the leg scanning group and six patients in the IPG alone group returned with venous thrombosis documented by objective testing. Not a single patient in either group suffered fatal pulmonary embolism. Remarkably, follow-up to 12 months was achieved in 911 of 912 patients; all 645 patients randomized were followed to 12 months.

The authors contend that the findings indicate that serial IPG alone is effective and it is therefore safe to withhold therapy in patients who remain negative by serial IPG. This latest experiment once again met the high methodological standards established by this group in their previous work; randomization, strict adherence to protocol, and complete follow-up were all achieved. The assessment in this study focused on the relevant aspects of patient outcome: initial treatment for venous thrombosis, recurrent venous thrombosis, pulmonary embolism, and death. Complications suffered as a result of anticoagulation are not detailed, but could only be greater in the IPG and leg scanning group in which more patients were anticoagulated. That no differences existed between the two groups on any of the major measures of outcome suggests that patients in whom leg scanning was administered suffered, as a group, greater inconvenience and risk associated with long-term anticoagulation, without reaping any benefit of decreased risk of recurrent thrombosis, pulmonary embolism, and death. Therefore, the authors' conclusions are warranted. This study is an excellent example of how the effect of a diagnostic test on outcome can be determined in the setting of a controlled trial.

Impedance Plethysmography: Economic Evaluation

In association with the clinical assessments discussed above, the same team of researchers (Hull et al., 1981) examined the economic effect of IPG and other diagnostic tests in the treatment of deep venous thrombosis. The main economic question was whether the extra resources consumed in diagnostic testing could be justified in terms of the potential savings in resources, because more accurate diagnosis would avert the need for costly therapy in some patients. The economic analysis was based on the results for 478 patients with clinically suspected venous thrombosis who were referred to the Hamilton District Thromboembolism

Programme and in whom the full range of diagnostic tests could be adequately performed. The clinical data enabled a comparison of costs of the possible alternative diagnostic strategies and their effectiveness in terms of cases correctly diagnosed.

For the purposes of the economic evaluation, six strategies were compared:

- clinical diagnosis (without the use of objective diagnostic tests)
- inpatient venography (the "gold standard" test)
- outpatient venography (a cheaper approach than inpatient venography, but available in only a few locations)
- IPG alone
- IPG plus leg scanning with ^{125}I-fibrinogen
- IPG plus outpatient venography if IPG is negative

According to clinical diagnosis alone, all 478 symptomatic patients would receive care and treatment in the hospital. The costs would include those of the anticoagulant drugs (intravenous heparin followed by warfarin sodium given orally), of hospitalization due to the extended inpatient stay and of monitoring therapy.

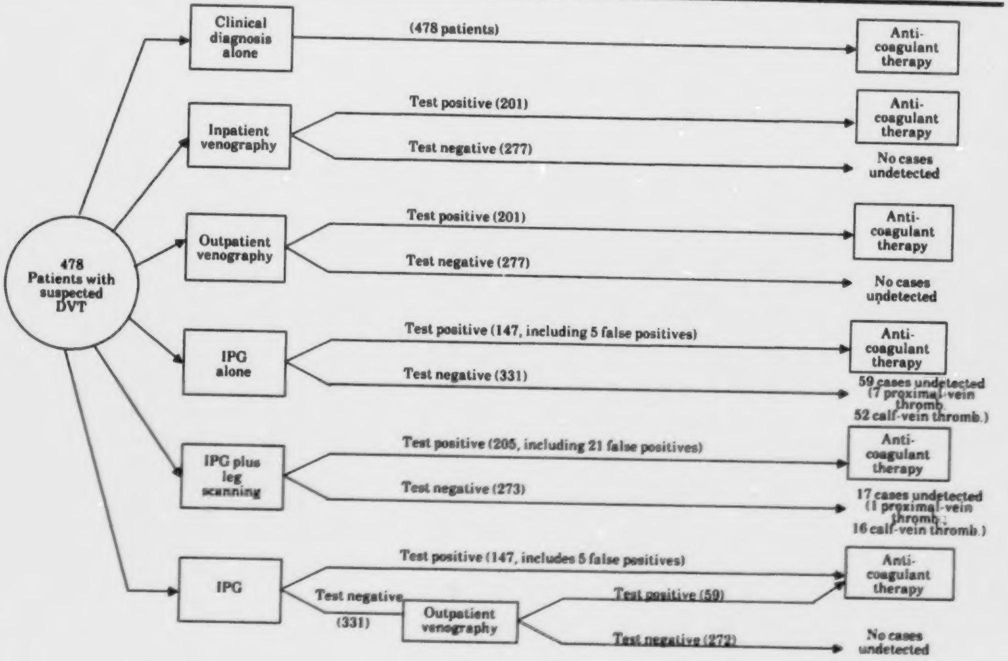
Venography, given as an inpatient or outpatient procedure, correctly "rules in" or "rules out" disease. Consequently, many fewer the patients (n = 201) would receive therapy, although other costs, as for performing the venograms, would be incurred. Also, venography is an invasive procedure with minor complications.

As outlined above, the non-invasive strategies (IPG alone and IPG plus leg scanning) are less accurate than venography. Therefore, although the costs of diagnostic testing would likely be lower than of venography, a number of cases of DVT would be missed. False positives would also result in patients receiving anticoagulant therapy unnecessarily.

The final approach, IPG followed by outpatient venography if IPG is negative, would combine the best of both worlds. It would lead to less invasive testing than venography alone, yet would still correctly identify all the cases of DVT. Though its higher diagnostic costs are slightly higher than IPG alone, this approach might be justified because of higher case finding, especially if undiagnosed DVT led to health problems in the future.

The six strategies are set out in diagrammatic form in Figure 9.1. The results are based on the clinical study carried out by Hull and colleagues (1981).

Figure 9.1
Diagnosis of DVT: Algorithm of Clinical Alternatives


478 Patients with suspected DVT
Clinical diagnosis alone
(478 patients)
Anti-coagulant therapy
Inpatient venography
Test positive (201)
Anti-coagulant therapy
Test negative (277)
No cases undetected
Outpatient venography
Test positive (201)
Anti-coagulant therapy
Test negative (277)
No cases undetected
IPG alone
Test positive (147, including 5 false positives)
Anti-coagulant therapy
Test negative (331)
59 cases undetected (7 proximal-vein thromb.; 52 calf-vein thromb.)
IPG plus leg scanning
Test positive (205, including 21 false positives)
Anti-coagulant therapy
Test negative (273)
17 cases undetected (1 proximal-vein thromb.; 16 calf-vein thromb.)
IPG
Test positive (147, includes 5 false positives)
Anti-coagulant therapy
Test negative (331)
Outpatient venography
Test positive (59)
Test negative (272)
No cases undetected

Application of the Methodological Guidelines

Ten guidelines for undertaking economic evaluations of new health care technologies were given in Chapter 7. In this section we examine the extent to which these were followed in the economic evaluation of IPG and other diagnostic tests in the diagnosis and treatment of DVT.

A service-level of approach was adopted by the researchers, because it was advantageous to build on the results of the clinical assessments that were already being carried out. The evaluation therefore relates only to this particular application of IPG and not to the total systems effect of the new technology. There is no evidence that IPG is used in other applications. Leg scanning has been proposed as a prophylactic measure in the prevention of fatal pulmonary embolism in high-risk surgical patients. Alert to this possibility, the same authors (Hull et al., 1982) have undertaken another study which shows that on economic grounds leg scanning is inferior to primary prophylaxis (subcutaneous administration of heparin in low doses).

The viewpoint for the economic evaluation of IPG was clearly identified as that of the Ministry of Health. The major cost items included the payments made by the ministry through the Ontario Health Insurance Plan (OHIP) and global budgets to hospitals. Although in some cases such a viewpoint would be unnecessarily restrictive, in this case it is likely that adoption of a wider viewpoint would have merely confirmed the results obtained. (That is, exclusion of, say, the production losses to the community due to hospitalization of patients is conservative for the hypothesis that IPG is economically worthwhile.)

A number of alternatives to IPG were considered. Clinical diagnosis was included since it was the common practice before the advent of objective testing for DVT. In addition, the main rival technologies were examined, the other main non-invasive test (leg scanning) and the "gold standard" test (venography). Finally, combinations of the various tests were considered, because this more accurately reflects clinical reality. That is, the normal practice is for a non-invasive test to be given first and then, if positive, confirmed by a more accurate, but possibly invasive, test.

The form of economic evaluation chosen was cost-effectiveness analysis, the alternatives being compared for the total costs incurred and the number of correct diagnoses obtained. That is, it was tacitly assumed that the objective was to diagnose correctly and to treat DVT.

The long-term clinical data on the consequences of missing DVT were not known at the time of the economic evaluation.

Had they been available, it might have been possible to undertake a more sophisticated cost-utility analysis. This would, in principle, have embraced a broader range of factors including life expectancy and the unpleasantness of the invasive tests, of the side effects of anticoagulant therapy, and of the pain of recurrent untreated DVT.

The range of costs and benefits considered was largely conditioned by the methodological choices outlined above. It was restricted to those costs and benefits to the Ministry of Health arising from the choice of diagnostic strategy. These included the costs of the diagnostic tests (technical components and physicians' fees) and the costs of anticoagulant therapy (drugs, laboratory tests to monitor therapy, physicians' fees, and hospitalization). As was mentioned earlier, other costs could have been included, such as those out-of-pocket expenses falling on patients and costs to the community in lost production. Omission of these was conservative as they are likely to be much higher for the alternatives to IPG, which require more hospitalization.

The main methodological issue in the estimation of costs and benefits was the calculation of hospitalization costs. The approach adopted was to assume a cost of $250 per day to reflect the portion of costs that remains approximately unaffected (such as general nursing, maintenance, meals, and housekeeping) by the special needs of patients with this particular diagnosis. A methodologically superior approach would have been to undertake a more detailed costing exercise within the hospital, with the objective of identifying more precisely the costs attributable to thromboembolism patients. (This was the approach adopted in the case study of neonatal intensive care described in Chapter 10.) The choice of approach depends on the importance of the particular estimate to the overall study results and the extent to which the results change if different estimates are assumed. (This issue is discussed in the context of this example in the section on sensitivity analysis, below.)

The economic evaluation of diagnostic tests in this field has recognized the distinction between average and marginal costs in two ways. First, in the presentation of the results of the study of diagnostic tests for DVT, an incremental cost-effectiveness analysis was performed. Consider the data in Table 9.1. Hull et al., (1981) determined the total costs and results (in case findings) of the alternative strategies for diagnosing DVT and provided the data for analyzing the incremental costs and effects of one diagnostic strategy over another. For example, IPG alone is the cheapest strategy but picks up fewer cases than a combined approach of IPG plus leg scanning. The incremental cost of the combined approach over IPG alone is $154,687 (in Canadian dollars) and 42 more cases are found; an incremental costs to effectiveness (C-E) ratio of inpatient venography (the commonly used "gold standard") over the combined approach is $13,852. Thus the price being paid for extra testing can be clearly identified and a judgement of its usefulness made, based on the clinical importance of missing some cases of DVT. (As already mentioned, the latest evidence suggests that the cases of calf DVT missed by IPG do not lead to adverse consequences.) Data on total costs per correct diagnosis are presented in Table 9.2.

TABLE 9.1

Total Costs and Results of the Alternative Strategies to Diagnose Deep-Vein Thrombosis

Diagnostic Approach	No. of Correct Diagnosis	Cost	
		Canadian $	American $
Impedance plethsmography alone	142	$ 395,359	$ 321,488
Combined impedance plethysmography and leg scanning	184	550,046	456,504
Outpatient venography	201	570,863	646,215
Inpatient venography	201	785,538	845,004
Clinical diagnosis	201	1,260,046	986,508
Impedance plethysmography plus outpatient venography if impedance plethysmography is negative	201	577,609	603,552

Source: Hull et al (1981), p. 1564.
Reprinted by permission of the *New England Journal of Medicine*.

Note: Canadian costs are given in 1980 Canadian dollars.

TABLE 9.2
Total Costs of the Alternative Strategies for Each Patient with Venous Thrombosis Correctly Managed

Diagnostic	Total Cost Per Patient with Correct Diagnosis and Treatment			
	Proximal-Vein or Calf-Vein Thrombosis		Proximal-Vein Thrombosis	
	Canadian $*	American $+	Canadian $*	American $+
Clinical diagnosis	$6,319	$4,908	$9,137	$7,098
Venography – Outpatient	2,840	3,215	4,107	4,649
Elective inpatient	3,908	4,204	5,651	6,080
Impedance plethysmography alone	2,784	2,264	2,995	2,436
Impedance plethysmography plus leg scanning	2,989	2,481	3,986	3,308
Impedance plethysmography plus venography	2,874	3,003	4,155	4,342

* Canadian dollar cost in Ontario, 1980 Canadian $.

+ United States dollar cost in an urban New England Hospital.

Source: Hull et al (1981), p. 1564.
Reprinted by permission of the *New England Journal of Medicine*.

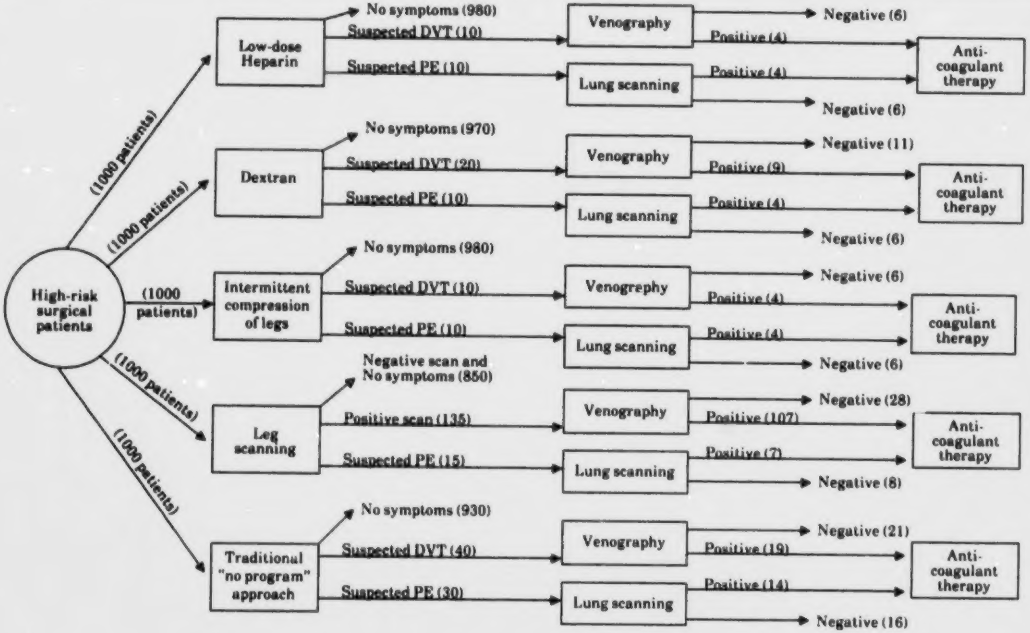
Second, another economic evaluation in this field, on prophylaxis for PE in high-risk surgical patients, makes the average-marginal distinction in another way (Hull et al., 1982). This study compared treatment of clinically suspected PE, as it occurred, with a range of prophylactic approaches, including leg scanning, intermittent pneumatic compression of the legs, and primary prophylaxis with low-dose anticoagulants – heparin or dextran (see Figure 9.2). A major finding was that a secondary prophylactic approach involving leg scanning, followed by venography for positive scans, was more costly than primary prophylaxis. The message here is that while leg scanning might form one component of a cost-effective approach in one clinical application, it may be cost ineffective in another. This is likely to be true of most new medical technologies.

The issue of differential timing of costs and benefits was not a major consideration in the economic evaluation of diagnostic strategies for DVT, as under all approaches the resources were committed in a fairly short period of time. The discounting of costs and benefits would need to be considered, however, if it were found that cases of DVT missed by IPG did lead to adverse clinical effects and resource use at a later date. Also, for a hospital not having some of the new diagnostic equipment, the costs of purchasing IPG or leg scanning machines would need to be considered. If it were more convenient, from the point of view of the analysis, to express these capital costs as an annual recurring amount over the life of the equipment, the differential timing method outlined in Chapter 7 could be used. (The same would be true if extra building works were necessary to make the new testing procedures available in some hospitals.)

Because the economic evaluation of diagnostic tests for DVT was based on good clinical assessment, much of the uncertainty in costs and benefits was removed. The major factors for which it was thought important to perform a sensitivity analysis were the prevalence of deep-vein thrombosis, the daily cost of hospitalization, the cost of anticoagulant therapy, and the costs of the diagnostic tests used. Fairly extreme ranges were used for those factors likely to change the results of the study. It was assumed, for example, that the daily cost of a hospital bed may be as low as 40 per cent of the original figure ($250 per day) assumed. The cost-effectiveness ranking of the diagnostic alternatives was not changed by varying the factors mentioned.

The major attribute of the presentation of results of the DVT study was the incremental cost-effectiveness analysis already mentioned. In addition, the authors examined the ranking of diagnostic alternatives in another setting, an urban teaching

Figure 9.2
Algorithm of Phrophylaxis for Pulmonary Embolism in High-Risk Surgical Patients


High-risk surgical patients
(1000 patients)
(1000 patients)
(1000 patients)
(1000 patients)
(1000 patients)
Low-dose Heparin
No symptoms (980)
Suspected DVT (10)
Suspected PE (10)
Venography
Negative (6)
Positive (4)
Lung scanning
Positive (4)
Negative (6)
Anti-coagulant therapy
Dextran
No symptoms (970)
Suspected DVT (20)
Suspected PE (10)
Venography
Negative (11)
Positive (9)
Lung scanning
Positive (4)
Negative (6)
Anti-coagulant therapy
Intermittent compression of legs
No symptoms (980)
Suspected DVT (10)
Suspected PE (10)
Venogrephy
Negative (6)
Positive (4)
Lung scanning
Positive (4)
Negative (6)
Anti-coagulant therapy
Leg scanning
Negative scan and No symptoms (850)
Positive scan (135)
Suspected PE (15)
Venography
Negative (28)
Positive (107)
Lung scanning
Positive (7)
Negative (8)
Anti-coagulant therapy
Traditional "no program" approach
No symptoms (930)
Suspected DVT (40)
Suspected PE (30)
Venography
Negative (21)
Positive (19)
Lung scanning
Positive (14)
Negative (16)
Anti-coagulant therapy

hospital in New England. It was found that the results were broadly similar, although strategies involving venography were relatively more expensive in the United States.

The general conclusions of the study were that clinical diagnosis alone was cost ineffective and that objective testing technologies were justified. The study was more equivocal with respect to the choice among objective testing strategies, as the long-term consequences of untreated calf DVT were not known at the time. However, the economic advantages of outpatient diagnostic approaches, using IPG, leg scanning, or venography were stressed.

The economic evaluation of diagnostic strategies for DVT (Hull et al., 1981) and associated economic studies in related clinical fields (such as Hull et al., 1982) show that, when appropriately conducted, such studies can assist in making rational decisions about the use of new diagnostic technologies. In particular, these studies highlight the advantages of integrating economic evaluation closely with clinical assessment.

As the diagnostic technologies in the field of venous thromboembolism develop, further economic studies will be possible. For example, new data on case findings in patients with clinically suspected pulmonary embolism (Hull et al.,1983) will make possible an economic evaluation of performing extra diagnostic tests (such as pulmonary angiography, ventilation lung scanning, and venography) on patients who have an abnormal perfusion lung scan. The appropriate diagnostic strategy in such patients is currently a source of debate (Robin, 1977; McNeil and Bettman, 1982). It is also possible that IPG may have a role in the diagnosis of those patients who are too sick to have a venogram or angiogram. Finally, new technologies, such as digital angiography, would need to be evaluated on clinical and economic grounds in this clinical field.

Conclusion

What conclusions should a policy maker draw given these clinical and economic assessments of IPG?

First, it is clear that the criteria outlined in our guidelines for technology assessment discussed in Chapters 4 through 7 have largely been met in the case of IPG. Although some of the methodological issues may not have been fully tackled, the quality of the clinical and economic evidence surpasses that available for most diagnostic technologies. The general picture is that of a safe, non-invasive technology that has the potential to lead to the provision of more effective and efficient health care.

Second, a number of specific strengths of the technology, and its evaluation, can be stressed.

1. It has been established by medical research that the condition concerned, deep-venous thrombosis, is linked to pulmonary embolism, a serious, life-threatening condition.
2. An effective therapy is available, which happens to be the same for both DVT and PE. The dosage levels of the anticoagulant therapy are well-documented and are based on clinical research.
3. Given the seriousness of the condition, it is likely that when physicians suspect PE but do not have an objective diagnostic test like IPG available, they would in the main elect to treat. Therefore, there is no question of extra diagnosis leading to an add-on in extra treatment. In fact, the economic evaluation of IPG shows the reverse. The extra diagnostic information, in ruling out DVT or PE, leads to a reduction in the amount of anticoagulant therapy, and consequent cost savings.
4. IPG is less invasive than the current "gold standard" test, venography. Thus it reduces the unpleasantness of investigation from the patient's point of view.
5. The costs and effects of IPG have been evaluated, alongside venography and other non-invasive tests, such as leg scanning. Its false positive rate is low, and much lower than that for leg scanning. Its false negative rate is high, however; approximately 29 per cent of cases of DVT were missed in the clinical study reported above. It turned out that the majority of the cases missed were calf DVT, which did not lead to serious clinical consequences later. IPG's sensitivity for proximal DVT is over 95 per cent; therefore most DVT's likely to be complicated by PE will be detected. The insensitivity of IPG to distal thrombosis does not lead to poor clinical outcomes; it also saves the cost and side effects of anticoagulant therapy in patients with undetected calf DVT.
6. The issue of expansion in the clinical indications for non-invasive testing has been addressed in the research on DVT and PE. It has been shown that use of leg scanning in high-risk surgical patients as a prophylactic measure is cost-ineffective.
7. IPG does not involve large set-up costs, as with some diagnostic technologies, such as nuclear magnetic resonance (NMR) imaging. The IPG equipment costs approximately $11,000 and can be operated by a technician with the appropriate training.

Finally, although the evaluations carried out to date have a number of strengths, some uncertainties inevitably remain. For example, most of the research on IPG has been carried out in a tertiary referral centre with a special interest in venous thrombosis. If IPG were to spread to other centres would the clinical caseload be the same? Would the technology achieve the same results in potentially less-skilled hands?

These factors will need to be carefully monitored as the technology diffuses into the health care system. Nevertheless, the evidence supports use of IPG as the primary investigation in patients suspected of having deep vein thrombosis and the dissemination of the technology for that use.

Nuclear Magnetic Resonance: Clinical Assessment

Rather than applying our guidelines to technologies that have already been evaluated, we will focus in our next case study on an emerging diagnostic technology that may have a substantial effect on imaging methods. Nuclear magnetic resonance is a complex and sophisticated new method of producing images of the human body. The possibilities of this new technology are impressive enough that, as the authors of a recent paper describing NMR imaging of the brain observed, "it is difficult to restrain the current enthusiasm and pressure for commercialization of the technology" (Brant-Zwadzki et al., 1983). Later in this book we will examine how such restraint might be encouraged; in the present discussion, however, we will concentrate on how our guidelines can be applied in planning the assessment of NMR imaging.

The mechanisms whereby conventional radiography, ultrasound, and nuclear scans produce their images is relatively easy to understand; NMR is considerably more complex, and a detailed explanation is beyond the scope of the present discussion. Suffice it to say that placing an object, or a living organism, in strong magnetic field (or fields) and manipulating appropriate radio-frequencies (by rapidly turning them on and off, for example) perturbs the atoms within the object or organism so that they will subsequently emit a signal. The signal can be monitored either to determine the chemical structure of the tissue or to produce a pictorial image. The discussion will be limited to the latter application, in which the concentration of hydrogen ions (generally proportional to the water concentration) in tissue is utilized to produce the desired images.

Several potential advantages of NMR over available imaging techniques are responsible for the enthusiasm it has generated. First, it does not involve the use of ionizing radiation; in fact, experiments to date have not turned up any major adverse effects. Second, although with computed tomography (CT) scanning one can only "slice" the body in one plane (and attempts to reformat the image in other planes have met with variable success), NMR allows cross-sectional views in any plane desired. Third, the great radio-density of bone leads to what are called bone artefacts on CT; tissues adjacent to bone may be distorted. Such artefacts are absent with NMR, because in NMR bone provides no signal. Finally, because NMR distinguishes between tissues according to water content, there is better contrast among different types of soft tissues than in other available techniques. The grey and white matter of the brain, for example, can be differentiated, and there is a possibility that normal and abnormal (such as malignant) tissues can be distinguished as a result of differences in tissue water concentration and chemical environment.

Given these advantages, it is plausible that NMR may add unique diagnostic information, leading to superior performance, in comparison to existing technologies. If, however, NMR is really safer than its major competitor, CT scanning, do we really require that it demonstrate superiority? Would equivalence not be sufficient? If demonstration of equivalence were all that was desired, there would be extremely important implications for the evaluation process, in that one would require only an examination of accuracy, rather than needing to study therapeutic effect and patient outcome. There are, however, a number of problems with NMR, not the least of which are the larger capital investment required for NMR than for CT and the decreased patient throughput that results from the increased time required for imaging. Other problems include the unique environmental requirements of the NMR (protection from radio-frequency pollution); the fact that NMR imaging can be disturbing for the patient both because of claustrophobia and because of the noise created by the pulsating radio-frequencies; and the fact that sick patients, or those with monitoring devices, cannot have NMR performed (because of problems with magnetic fields and with observing patients during the procedure). Finally, there is always the possibility that NMR has adverse effects of which we are presently ignorant. It is crucial, therefore, that NMR be carefully investigated with a view to determining not only its accuracy, but whether it alters therapy and outcome.

The initial investigations of NMR should be of the sort conducted to date; experimenting with the technique to determine how to obtain the best images, applying it to different tissues to see what information it can provide, and obtaining an initial impression of how it compares with available technologies. This exploratory stage is particularly important for a technology such as NMR, in which manipulation of technical variables may lead to more informative imaging.

NMR has reached the stage where we have begun to understand many of its characteristics and have a good idea of its potential applications; pressure is mounting for its application in routine service. These facts argue for rapidly commencing the sort of definitive studies that will allow rational decisions about the role of the new device.

We have previously described how the approach to assessment of diagnostic technologies must be based on asking the right questions. This means choosing a particular presenting problem, examining the effects of the technology on how the health worker solves the problem, and studying the ultimate effect on the patient. The first question in evaluating NMR is what problems should be selected. Because the technique could be applied to many problems in which imaging is required, the choice is broad. Therefore, we must establish criteria for choosing a suitable problem. Some of the factors influencing the choice are as follows:

1. *Frequency*. The problem chosen must be common enough that it will not take an inordinate amount of time to collect a sufficient number of cases to make valid statements about the effect of NMR. This is an especially important consideration when one is dealing with a technology in which the pressures for rapid dissemination are intense. It would be very interesting to find out, for example, if whole body NMR scanning could help us sort out the difficult cases involving patients with fever of unknown origin; but by the time we knew the answer, the issue of its dissemination would have been long settled. Also, selection of frequent cases concentrates on areas where the potential costs and benefits of applying the new technology are large.
2. *Availability of a gold standard*. Assessing the accuracy of a new technology is far easier if there is a gold standard available. In the absence of a gold standard, we must resort to techniques of construct validation or infer the accuracy of a technique from its ability to alter patient outcome. The gold standard result must also be readily available, as we would not want to wait for years to find out whether the new technique is really accurate.
3. *Effect on patient management*. There is little point in evaluating the effect of NMR on a problem in which the

results will not affect management, for in such situations there is little opportunity for the new technology to provide benefit (and thus, little reason to institute it as a clinical procedure). It has been found, for example, that in multiple sclerosis, a degenerative crippling disease of the central nervous system for which there is no effective treatment, NMR has a remarkable capacity to show the demyelinating lesions in the brain. Yet, because the extent of the lesions will not influence management, evaluating the effect of NMR on the problem of assessing the progression of multiple sclerosis would be inappropriate. It could be argued that application of the technology will have tangible benefits in terms of physician confidence or in patients having a better idea of where they stand. It seems equally plausible, however, that documenting the inexorable progression of the disease may have adverse consequences on the patient's psychological state. At any rate, it would be incumbent on those who advocate use of NMR for this purpose to demonstrate its benefit.

It should be noted that this argument refers only to the clinical uses of NMR; it may still be that understanding of multiple sclerosis will be furthered by use of NMR in research. Its use in research, a special case, would obviously be far less expensive than widespread clinical dissemination and would require separate justification on the basis of scientific promise and ethical criteria.

A final point is what might occur if NMR was shown to be of benefit for other indications and was installed in the hospital. Would one not expect the curious neurologist to seek documentation of the demyelinating lesions in multiple sclerosis patients under his or her care? One certainly would, and this raises the problem of limiting the use of technologies to indications for which there is evidence of benefit.

4. *Change in management and change in outcome.* A problem is especially attractive for assessing a diagnostic technology (here, NMR) whenever there is a close and well-established link between changes in management and changes in outcome. The reason is that it is generally far easier to document changes in management than changes in outcome, and if one can make the assumption that the latter will follow from the former, the difficulties of the assessment process will be ameliorated. To return to the issue of deciding on surgical resectability of carcinoma of the lung, for example, if one could demonstrate that

application of NMR obviated the need for mediastinoscopy (a procedure with known morbidity), benefit could be considered to have been established.

5. *Deficiencies of existing methods.* Head trauma is a common problem in which management decisions are both affected by the results of imaging and have a known effect on outcome. It would not be a particularly wise choice of a problem in which to compare CT and NMR, however, because the performance of CT is already remarkable. In general, if the new technology may improve on existing methods, one would choose a patient presentation in which current methods are unsatisfactory. If one were to pick an area in which the existing technology is already completely satisfactory, the outcome would be a foregone conclusion; the new technology could at best match the results of existing techniques, and if, as in this case, the new technology is more expensive, one would end up rejecting its use. This argument would not apply if NMR were both less hazardous and less expensive than CT; demonstration of equivalence between the two techniques in management of head trauma might then become a question of very high priority.

No single problem is likely to meet all these criteria in a completely satisfactory manner; therefore, one must trade off among them. If, for example, the outcome of a particular problem is radically affected by choice of management, and the new technique promises markedly to increase the accuracy of diagnosis, one may decide to study the problem despite its not being as common as one would like. Another example is the one we have chosen to use to illustrate the application of our guidelines to NMR: suspected lesions of the brain and spinal cord. If broadly defined, this problem is common; a definitive answer is available within a relatively short period of time either through invasive tests or through follow-up of patients in whom no evidence of a brain or spinal cord lesion appears; test results will alter management; and existing diagnostic methods are not completely adequate. For some underlying causes of suspected space-occupying lesion, such as subdural hematomas or bleeding into the posterior fossa, the link between management and outcome is clear; for others, such as malignant gliomas, it is not.

We shall describe an approach to determining the usefulness of NMR head scanning in which the accuracy, therapeutic effect, and diagnostic effect of the technology are assessed in a single study. This is certainly not the only way the investigation could be undertaken, and indeed there are many satisfactory approaches.

Objective: Asking the Appropriate Question

The objective of the study would be to assess the accuracy, therapeutic effect, and effect on outcome of NMR scanning in patients with suspected intracranial space-occupying lesions, relative to conventional assessment with CT scanning.

We previously pointed out that it would make little sense to ask whether NMR scanning of the head was any good; one needs to define the question in terms of a specific clinical problem. The issue arises whether suspected intracranial space-occupying lesion is sufficiently specific. There are theoretical reasons to think that NMR may be particularly good at detecting lesions of the posterior fossa, arteriovenous malformations, subdural hematomas, pituitary adenomas, and spinal cord lesions. This might be the case, and yet because these conditions will be found in only a small proportion of those referred for imaging with suspected lesions of the brain or spinal cord, NMR might appear no better than CT scan in the overall analysis. Therefore, it might be better to define the question more narrowly (suspected posterior fossa lesion, suspected spinal cord lesion, etc.). Unfortunately, what one would gain in the precision of the question with this approach, one would lose in the additional time required to accrue enough patients to obtain a valid answer. One solution is to define the question broadly, specify a number of sub-hypotheses (such as that NMR will have a great advantage over CT scanning in spinal cord lesions), and at the end of the study analyze the data both in terms of the overall question and the secondary hypotheses.

Study Design

Patients with suspected intracranial space-occupying lesion would be randomized to receive NMR or CT scanning as the clinical test. Initially, all scans would be read blind to clinical information; the readers would then be provided with the history, physical examination, and other tests results, and the interpretation would be revised if appropriate. The interpretation would be recorded on standardized forms that would allow subsequent construction of receiver operating characteristics (ROC) curves. Having readers provide interpretations both before and after access to clinical information would allow determination of the accuracy of the scans when interpreted in isolation and in the clinical context in which they would ultimately be used. Patients in both groups would also receive the alternative diagnostic technology, which would not be released to clinicians taking care of the patients. Performing both procedures on all subjects would substantially increase the power of the study to assess the relative accuracy of CT and NMR scanning.

The intraobserver reliability could be determined by having all readers review the scans in random order some months after the initial reading and by comparing their interpretations to those initially provided. Interobserver reliability could be tested by having multiple readers review the same scans. These reliability studies could be conducted independently of the clinical schedule and need not interfere with a prompt report being provided for the attending physician.

According to the randomization schedule, clinicians would receive the results of either the NMR or the CT scans and would manage the patient on the basis of the information provided. Prior to receiving these results, they would further investigate and manage the patient if the scan were not available. On receipt of scan results, physicians would be asked to revise their list of likely diagnoses and to construct a new management plan. Patients would be followed over the succeeding months to determine what further investigations and treatment were in fact undertaken, the morbidity (including days off work, days confined to bed, days in hospital) and mortality they suffered, and their quality of life.

Patient Selection

To ensure that the results may be generalized, one would want to enrol consecutive patients referred to the department of radiology for head scanning because of suspected intracranial space-occupying lesion. Considerations for generalizing may be tempered if, for example, a large number of patients suffering from headache but without objective neurological findings were referred for imaging. It is known that CT scanning in such individuals is almost always negative, and one might waste resources if such individuals were included in the sample. A solution would be to exclude patients referred because of headache, but without neurological findings.

Measurement of Outcome

Accuracy and diagnostic effect. Adequate reliability is necessary for valid results; intraobserver and interobserver reliability would be assessed by the kappa statistic for both CT and NMR scans.

The gold standards for determining accuracy would include the results of biopsies, surgical procedures, and autopsies, and prolonged symptom-free follow-up. For each subject CT and NMR results would be available, both before and after clinical information was provided for the readers. ROC curves for each of the four conditions would be constructed and compared. Additional information about the diagnostic effect of the scans would be obtained from analysis of the manner in which scan results changed the physicians' estimates of the diagnostic probabilities.

Effect on the health care provider. A questionnaire could be administered to the attending physicians asking them to rate their own anxiety about the patient's condition and likely outcome, about the diagnosis, and about whether the optimal management is being delivered. They would be asked these questions before obtaining the results of the scan (under the assumption that it was unavailable), after the results were back, and several times thereafter. Physician responses would provide data concerning the reassurance value of NMR and CT scanning for the physician.

Therapeutic effect. Estimates of therapeutic effect would be obtained by comparing physicians' plans before and after scanning and by comparing physicians' plans with what was actually done in the clinical situation. NMR and CT scanning could be compared according to such variables as the number of times physicians changed their plan after receiving the results and the nature of the revisions. The RCT design would allow a definitive demonstration of therapeutic effect by comparing what was actually done in the two groups. An analysis of the treatments given would be strengthened by stratifying according to final diagnosis. One would examine patients who ended up being normal, those with various types of brain tumour, aneurysms, arteriovenous malformations, subdural hematomas, and so on, and examine the nature, time of commencement, and duration of treatment received by patients in the NMR and CT groups.

Patient outcome. Despite the widespread impression that introduction of new technologies like CT scanning have been of great benefit to patients, and that the advent of NMR is likely to have similar effects, the need for objective assessment of patient outcome is highlighted by the difficulty one has even imagining plausible situations in which the new technology will influence clinically relevant endpoints. There may be situations in which mortality is affected, such as a subdural hematoma missed by CT scanning that NMR would have picked up. Although mortality should be measured, situations like the above will be sufficiently rare that it is very unlikely the mortality would differ in the two groups. Morbidity may be influenced by randomization to NMR or CT scan; for example, the complications of subsequent angiography may be avoided if NMR proves to show posterior fossa lesions extremely well. In addition, serial measures of quality of life would be obtained and might show differences resulting from, say, more expeditious removal of a subdural hematoma, early treatment of a pituitary adenoma, or avoiding a craniotomy because of ability of NMR to distinguish between a tumour and a resolving hematoma.

Adverse effects. Investigators would be alert to the possibility of psychological or physical toxicity from either of the two procedures. A questionnaire would be administered to patients following the procedure inquiring how unpleasant or uncomfortable it is for them. Results would probably show NMR to be more unpleasant than the CT scan because of claustrophobia and noise from the rapid oscillations in the radio-frequencies. One source of adverse effects from the CT scan would be the idiosyncratic reactions sometimes seen to the use of contrast to enhance the image; another would be the known effects of cumulative exposure to radiation on tumour induction. Use of paramagnetic substances as contrast enhancing agents for NMR is presently under investigation, and use of these substances may be a source of side effects of the procedure in the future. The investigation we have described could not deal with rare or delayed side effects; therefore, in addition to building an assessment of adverse effects into the RCT, one would require establishment of a registry to monitor the occurrence of unusual or long-term problems.

The challenges of the sort of evaluation we have described are obvious. We hope at this point it is clear, however, that rigorous assessment is necessary before we can be confident that dissemination of new technologies like NMR actually does any good at all.

Nuclear Magnetic Resonance: Economic Evaluation

The high set-up costs of NMR help to focus attention on its economic evaluation. The current purchase price of an NMR machine is between $1 million and $1.5 million, depending upon the magnetic field strength. The higher-strength machines, which may offer better imaging potential, use more expensive, super-conducting magnets and require more shielding. In addition, the capital building costs of NMR installations must be taken into account; the higher-strength machines may require a free-standing building on the hospital premises to avoid interference with other activities. Specially designed non-ferrous structures are available, costing from $1.5 million to $2.2 million. Another aspect of the set-up costs of NMR is the necessity of employing a medical physicist, if one is not already available. This is because of the fine tuning required to optimize image quality. Currently it is felt that NMR will never become as "automatic" as CT scanning, even in its post-development phase.

An economic evaluation would seek to answer the question: If NMR does some good, is the technology worth having when compared with the other uses to which the same scarce resources could be put? It would therefore logically follow, and build on, the clinical assessment of the new technology. A particular concern for the economic evaluation is that CT scanning, a technology very similar to NMR in its clinical application, was widely disseminated before the kinds of assessments discussed earlier in this chapter were carried out. (This concern will be addressed later in Chapter 11 on the policy for the evaluation and dissemination of new technologies.) In considering guidelines for the economic evaluation of NMR, however, it has to be recognized that economic studies may need to be designed in such a way that they can assist policy makers in this field, even though some of the data on the clinical assessment of NMR may not yet be available.

Application of the Methodological Guidelines

Ten guidelines for undertaking economic evaluation of new health care technologies were given in Chapter 7. In this section we examine how these might be followed in the economic evaluation of NMR.

The choice of level of approach in evaluating NMR is far from clear. It would be quite legitimate, for example, to adopt a service-level of approach, building on controlled clinical trials like the one for scanning patients with suspected intracranial space-occupying lesion, described in detail in the last section. The objective would be to assess NMR, compared with the main alternative diagnostic technology (CT in this case), in terms of the differences in the costs of making a diagnosis, the additional costs (or, indeed, savings) resulting from any differences in patient management, and the benefits resulting from any differences in final patient outcome. This approach to economic evaluation should certainly be contemplated for those NMR applications that are being evaluated from a clinical viewpoint. That is not to say that an economic component should be built into every clinical trial. In part this depends on whether the trial is concerned with assessing NMR in a service application or whether it is concerned mainly with answering a biological question, and the importance, in economic terms, of the particular NMR application. (The issue of incorporating economic analysis into clinical trials is discussed more fully in Drummond and Stoddart, 1984.)

A major problem with the micro approach to the economic evaluation (or to the clinical evaluation) of a diagnostic technology like NMR is that there are many potential applications in a wide range of medical specialities, including internal medicine, surgery, neurology, and oncology. It is unlikely that the clinical advantages of NMR will be rigorously demonstrated for all applications when the decisions to invest in NMR installations are being made. Therefore, it may also be pertinent to evaluate NMR at the level of the installation. This approach would be particularly valid in situations where it is difficult to control the uses to which the NMR machine, once installed, is put. Typically, one might expect that the machine would be used for a wide range of applications, some for which clinical effectiveness had been demonstrated, others for which effectiveness is more uncertain. This more system-level approach to evaluation would therefore seek to compare the total additional costs of using NMR in all its applications (effective and ineffective) with the total benefits, in terms of improved patient outcome or other desirable attributes.

As was indicated in Chapter 7, both evaluations are likely to be useful, depending upon the question asked and the stage in the diffusion of NMR. The system-level approach may be particularly important in the early stages, when clinical research into the various applications for NMR is being carried out. The system-level approach would have more prominence when decisions are being made about installation of NMR machines at tertiary or secondary care facilities. A related aspect is the issue of optimal facilities planning; that is, deciding upon the total number of installations, their size, and their location. This topic was dealt with in Chapter 8.

As was indicated in Chapter 7, the most relevant viewpoint for the economic evalution is that of society. Under this viewpoint all the costs and benefits of NMR would be considered, regardless of whether they fall on the government, the hospital, patients, or the community at large. (In the later parts of this section, it will be assumed that a societal viewpoint is being adopted, since it is the most comprehensive.)

It might also be worthwhile to consider NMR from two other viewpoints. First, from the hospital's viewpoint, the acquisition of an NMR machine may provide an opportunity to attract extra revenue, if it increases the number of diagnostic procedures for which health insurance plan billings can be made. Because tests referred in can be billed separately in most Canadian provinces, it will therefore be of interest whether NMR increases the number of billings. The effect of the installation of an NMR on a hospital's total budget is also relevant.

Second, from the government's viewpoint, NMR may be an add-on. This is likely to depend mainly on whether it becomes an adjunct to or replacement for other diagnostic tests and the relative cost of NMR compared with those tests. It will also depend on whether, on balance, NMR leads to additional therapy being undertaken or whether, conversely, more accurate diagnosis averts the need for costly therapy. Of course, the government can influence the effect of NMR on its own costs through its policy on reimbursement for NMR procedures. For example, once approved on safety grounds, all NMR investigations could be reimbursable. Alternatively, reimbursement could be restricted to those investigations for which NMR has been shown to be either effective or efficient. Furthermore, the government may not allow reimbursement for both NMR and CT scanning for the same clinical indication on the same patients unless this is of proven effectiveness. Obviously, such policies would have a number of advantages and disadvantages; they are discussed further in Chapter 11.

Turning to the issue of selection of alternatives for evaluation, it has already been pointed out that NMR has a potentially wide range of applications. It might be useful, when planning evaluations at the service-level, to think of these as falling into two broad categories: (1) those where NMR is essentially a replacement for existing diagnostic technologies and (2) those where NMR offers new diagnostic capability.

In the former case it would be most relevant to compare NMR with the technology currently in use; in many applications this may be CT scanning. In the latter case it would be relevant to compare NMR with a "baseline" alternative of clinical diagnosis alone. That is not to say that clinical diagnosis implies "doing nothing" since, in the absence of objective diagnostic tests, unnecessary therapy may be given, sometimes at high costs. (This was the situation in the economic evaluation of IPG, as discussed earlier.)

Obviously, the potential application for NMR may not fall neatly into these two categories. First, it may be that although

NMR is a replacement for an existing test, it may also offer new capability if it gives a more accurate diagnosis. Second, the existing diagnostic test, CT scanning, may not itself have been evaluated in accordance with the guidelines for clinical assessment described earlier. In this case it might be relevant to broaden the analysis to consider the clinical diagnosis option too. Third, there may be a number of choices within the diagnostic modalities being assessed, such as the number of views, or "slices", of the patient obtained. Studies of the wisdom of taking multiple views of the patient have been conducted in other branches of radiology (Cockshott, Jenkins, and Pui, 1983). Finally, there may be a number of situations where NMR is proposed for use in conjunction with existing diagnostic tests. Here the main question for evaluation would be whether the benefits of the extra information gained by adding NMR justifies the additional costs.

Although the range of potential applications for NMR is still expanding, it is already possible to identify situations where NMR might be compared either with an existing diagnostic technology or with clinical diagnosis alone.

In the system-level approach to evaluation, at the level of the NMR installation as a whole, the implicit alternative would be the pre-NMR utilization of diagnostic technologies in the given locality. When these include CT scanning, it is conceivable that if some investigations are diverted to NMR the spare capacity on the CT scanner may be filled by an expansion in the indications for CT scanning on other patients. This may or may not be desirable, depending on whether these extra investigations yield benefits for the patients concerned. The literature is somewhat confused on this point (Larson, Omenn, and Lewis, 1980; McNeil et al., 1981), but it is clear that the installation of NMR should not be considered without also considering the implications for the use of existing diagnostic technologies in each locality.

The selection of the form of economic evaluation is likely to depend crucially on the particular application of NMR being evaluated. If the application is broadly in the "replacement" category, then cost minimization may suffice. That is, the costs of two ways of meeting the same diagnostic objective would be assessed. It is hard, however, to envisage NMR being of lower cost than current technologies for many such applications, unless greater diagnostic accuracy averts the need for costly therapy. Therefore the economic attractiveness of NMR is likely to be stronger in applications where it offers new capability and where the benefits from this justify the higher costs. In such situations the appropriate form of economic evaluation is likely to be cost-effectiveness, cost-utility, or cost-benefit analysis. The results might be expressed, say, in incremental cost per year of life gained (or quality-adjusted life-years gained) from employing NMR instead of existing approaches.

The range of costs and benefits considered would depend on the question addressed and the viewpoint adopted. The general categories of costs important in the economic evaluation of an NMR installation are those for additional capital (NMR machine, building, alterations), additional staff (such as a radiologist, two technicians, a medical physicist), staff training, maintenance, consumables, and any extra clinical procedures arising from diagnosis by NMR. General categories of benefits are savings in existing tests that would be averted by use of NMR (for example, staff time, consumables, repairs and maintenance; plus, in the longer term, reduced replacements of equipment), savings owing to clinical procedures being averted following NMR diagnosis, and benefits to patients from the reassurance given,[1] the reduced invasiveness of NMR, and any effective therapy given as a result of NMR.

A number of issues might arise in the estimation of costs and benefits and in recognizing the distinction between marginal and average costs. First, if the service-level approach to evaluation is adopted and the costs of using NMR in a particular application are required, it will be necessaary to measure the time taken to investigate the patient (to compare with that of the existing diagnostic technology). This is best ascertained by a study of the throughput of patients in the room assigned to NMR using standard activity sampling techniques. Although NMR machines are still in the development stage, current indications are that the time required to investigate patients with NMR will be somewhat longer than with CT. Obviously, patient throughput is a key factor in determining the economic attractiveness of NMR, given the large set-up costs. Second, as was pointed out in Chapter 7, the marginal savings from reduction in other diagnostic tests and hospitalization of patients may be lower than the average costs.

Consideration of the differential timing of costs and benefits would be very relevant for NMR. First, the high set-up costs mean that a higher proportion of funds are being committed earlier than they would be if existing diagnostic procedures were continued. A common way of handling this would be to convert the initial capital outlay, in equipment and buildings, to an equivalent annual cost over their useful life using an agreed discount rate. (This procedure was discussed in Chapter 7.) An expenditure of $1.5 million on an NMR machine, for example, would represent, over five years at a discount rate of 5 per cent, an equivalent annual cost of $346,500. (This is higher than one-fifth of the purchase price, the figure that would be used in a simple straight-line depreciation of the equipment.)

Other ways in which the issue of timing of costs and benefits may be important in the evaluation of NMR include the fact that use of the technology may postpone or bring forward certain types of therapeutic intervention. Also, the benefits – avoidance of certain types of cancer through avoiding the use of ionizing radiation – will occur mainly in the future.

The major factors likely to give rise to uncertainty in costs and benefits include, first, the possibility of technological change, which might affect both the useful clinical life of the equipment and its purchase price, and second, lack of evidence on the benefits of improved patient outcome from using NMR. Trials of clinical effectiveness of diagnostic technologies are hard to mount and are unlikely to be available for all possible indications. This uncertainty will probably necessitate a sensitivity analysis using different assumptions. When the lack of knowledge is remediable, the carrying out of additional research should also be considered.

In presenting the results of any economic evaluation of NMR, it should be recognized that a number of uncertainties may remain and that the evaluation is an aid to decision making and not a substitute for it. Also there are likely to be differences between locations, in clinical caseload and availability of other diagnostic technologies, that may affect costs and benefits. Finally, other important factors, such as allowing equal access to diagnostic services by geographical area, may also have a bearing on whether or not an NMR machine should be installed in a given locality.

Notes

1. The value of reassurance to patients is difficult to assess but should not be overlooked. To include this category of benefit, one may have to go to a more sophisticated form of evaluation, such as cost-utility analysis or cost-benefit analysis, as mentioned in Chapter 7. Alternatively, one might simply list reassurance as a category of benefit and leave it to the decision maker to decide whether it justified any higher cost of NMR.

References

Brant-Zwadzki, M.; Davis, P.L.; Crooks, L.E.; et al. 1983. "NMR Demonstration of Cerebral Abnormalities: Comparison with CT." *American Journal of Radiology* 140: 847-854.

Drummond, M.F., and Stoddart, G.L. 1984. "Economic Analysis and Clinical Trials." *Controlled Clinical Trials* 5: 115-128.

Cockshott, W.P.; Jenkins, J.K.; Pui, M. 1983. "Limiting the Use of Routine Radiography for Acute Ankle Injuries." *Canadian Medical Association Journal* 129: 129-131.

Hull, R., et al. 1983. "Pulmonary Angiography, Ventilation Lung Scanning, and Venography for Clinically Suspected Pulmonary Embolism with Abnormal Perfusion Lung Scan." *Annals of Internal Medicine* 98(6): 891-899.

Hull, R.; Hirsh, J.; Sackett, D.L. 1979. "The Value of Adding Impedance Plethysmography to ^{125}I-fibrinogen Leg Scanning for the Detection of Deep Venous Thrombosis in High Risk Surgical Patients: A Comparative Study between Patients Undergoing General Surgery and Hip Surgery." *Thrombosis Research* 15: 227-234.

Hull, R.; Hirsh, J.; Sackett, D.L.; et al. 1977. "Combined Use of Leg Scanning: An Impedance Plethysmography in Suspected Venous Thrombosis." *New England Journal of Medicine* 296: 1497-1500.

Hull, R.; Hirsh, J.; Sackett, D.L.; Stoddart, G.L. 1981. "Cost Effectiveness of Clinical Diagnosis, Venography and Non-invasive Testing in Patients with Symptomatic Deep-Vein Thrombosis." *New England Journal of Medicine* 304: 1561-1567.

Hull, R.; Hirsh, J.; Sackett, D.L.; Stoddart, G.L. 1982. "Cost-Effectiveness of Primary and Secondary Prevention of Fatal Pulmonary Embolism in High-Risk Surgical Patients." *Canadian Medical Association Journal* 127: 990-995.

Hull, R.; Taylor, D.W.; Hirsh, J.; et al. 1978. "Impedance Plethysmography: The Relationship between Venous Filling and Sensitivity and Specificity for Proximal Vein Thrombosis." *Circulation* 58: 898-902.

Hull, R.; van Aken, W.G.; Hirsh, J.; et al. 1976. "Impedance Plethysmography Using the Occlusive Cuff Technique in the Diagnosis of Venous Thrombosis." *Circulation* 53: 696-700.

Johnston, K.W.; Kakkar, V.G.; Stimbler, J.J. 1974. "A Simple Method for Detecting Deep Venous Thrombosis: An Improved Electrical Impedance Technique." *American Journal of Surgery* 127: 349-352.

Larson, E.B.; Omenn, G.S.; Lewis, H. 1980. "Diagnostic Evaluation of Headache: Impact of Computerized Tomography and Cost-Effectiveness." *Journal of the American Medical Association.* 243(4): 359-364.

McNeil, B.J., and Bettman, M.A. 1982. "The Diagnosis of Pulmonary Embolism." In *Haemostasis and Thrombosis*, ed. R.W. Coleman et al. Philadelphia: Lippincott.

McNeil, J.; Hanley, J.; Funkenstein, H.H.; Rambaughm, C. 1981. "Utilization of Computed Tomography of the Head in a Tertiary Care Hospital." *Radiology* 139: 113-118.

Moser, K.N.; Brach, B.; Dolan, G.F. 1977. "Clinically Suspected Deep Venous Thrombosis of the Lower Extremities: A Comparison of Venography Impedance, Impedance Plethysmography and Radial Labelled Fibrinogen." *Journal of the American Medical Association* 237: 2195.

Richards, K.L.; Armstrong, J.D.; Tikoff, G.; et al. 1976. "Noninvasive Diagnosis of Deep Venous Thrombosis." *Archives in Internal Medicine* 136: 1091-1096.

Robin, E.D. 1977. "Overdiagnosis and Overtreatment of Pulmonary Embolism: The Emperor May Have No Clothes." *Annals of Internal Medicine* 87: 775-781.

Wheeler, H.B. 1973. "Impedance Testing for Venous Thrombosis." *Archives of Surgery* 106: 762.

Wheeler, H.B.; O'Donnell, J.A.; Anderson, F.A.; et al. 1974. "Occlusive Impedance Phlebography: Diagnostic Procedure for Venous Thrombosis and Pulmonary Embolism." *Progress in Cardiovascular Disease* 17: 119-205.

Chapter 10:

Clinical and Economic Evaluation of Therapeutic Technologies: The Case of Neonatal Intensive Care Programs

---

Kathryn Bennett & David Feeny

Clinical Effectiveness

This section illustrates the use of the guidelines for assessing therapeutic technologies by applying them to currently existing evidence on the effectiveness of neonatal intensive care (NIC) programs. The efficacy of these programs is discussed in the following section.

Over the past two decades numerous advances in the care of the newborn have been made. With these advances has come the establishment of special NIC units and regionalized NIC programs.

Neonatal intensive care programs involve the identification of high-risk neonates (infants less than twenty-eight days of age) and provision of appropriate care to ensure the best possible neonatal outcome. The services available within NIC programs require specially equipped units and laboratories staffed by specially trained personnel. The intensity and sophistication of this care has been defined in three levels: level 1, normal or low risk; level 2, moderate risk; level 3, high risk (Committee on Perinatal Health, 1976). A substantial number of obstetrical services offer level 2 or 3 care at this time. This section uses the guidelines for evaluating therapeutic technologies described in Chapter 5 to review the evidence on the efficacy of NIC programs and their effect on the community.

The first guideline asks whether the appropriate question was framed. This requires a clear specification of "good for doing what, to whom, and compared with what." For NIC, good for doing what refers to decreased mortality and improved functional status and quality of life among survivors; compared with what refers to no NIC or routine care. The specification of "to whom" should include a clear definition of the target population. Because NIC programs are usually intended to serve all infants born in a defined geographic region, the appropriate question will be framed in these terms. In addition, because neonatal risk varies with birth weight, the question should specify the birth weight of the population.

The recent evaluation of NIC reported by Boyle and colleagues (1983) illustrates this guideline. The authors clearly address each of these issues at the outset of the report in their statement of objectives.

The second guideline asks whether the design of the study was sufficiently rigorous for making policy decisions. Many of the specific subcomponents or procedures that make up NIC have been shown to be efficacious (Sinclair et al., 1981) in randomized trials. Any conclusions about the impact of the NIC package based on these studies require inferences, however, about the effects of the combination of subcomponents.

One randomized trial of the NIC package has been reported (Kitchen and Campbell, 1971). This trial evaluated the effectiveness of NIC compared with routine care in reducing neonatal mortality among infants weighing 1000 to 1500 grams at birth. A 49 per cent risk reduction for mortality was observed in the intensive care group. Unfortunately, because only infants born at the hospital offering NIC were included in the study, this limits the inferences that may be drawn about the effectiveness of NIC programs responsible for the care of all infants born in a defined geographic area. Evaluations based on only the outcomes of those who reach hospital may overestimate the program's effectiveness in the population served because not all babies will reach the facility. A bias would occur, for example, if the sickest babies die before they reach the NIC unit or are judged unlikely to survive the trip and are therefore not transported.

Despite the foregoing evidence of the therapeutic potential of components of NIC and the NIC package, the actual magnitude of effectiveness in the community should be determined from population-based studies of NIC programs. In the absence of a population-based randomized trial of NIC, such information is available from two sources: before-after studies comparing neonatal mortality and morbidity in two periods bracketing the introduction of NIC and descriptive studies of mortality and morbidity trends.

Six such studies of very low birth weight (VLBW) infants have recently been reviewed (Sinclair, 1982). All of the studies showed an important decrease in mortality associated with neonatal intensive care. The recent before-after study (Boyle et al., 1983) provides information on the magnitude of impact of NIC in a defined population. In this study infants born to residents in the Hamilton-Wentworth region in Ontario before (1964-1969) and after (1973-1977) the introduction of NIC were followed and their mortality and morbidity outcomes compared. Comparisons were made from birth to the time of hospital discharge and were projected to age 15 and the expected time of death.

A major problem with data from before-after studies is the lack of a concurrent control group, since the contribution to the mortality decrease of factors other than NIC that may have differed between comparison groups is open to question. Such factors relate to non-comparability of the comparison groups due to differing characteristics of the mothers and infants studied and the effect of co-intervention, such as improvement in antenatal and perinatal care. Co-interventions are discussed below under intervention bias (Budetti et al, 1981; Sinclair et al., 1981; Sinclair, 1982; Paneth et al., 1982a, 1982b; Shapiro et al., 1983; Bloom, 1984).

Several reports have shown that changes in the birth weight distribution account for less than 15 per cent of the observed reduction in mortality (Budetti et al., 1981; Williams and Chen, 1982; Kleinman et al., 1978; Lee et al., 1980). Analyses that control for birth weight consistently show birth weight specific declines in mortality. Investigations (within birth weight strata) of differences in maternal risk factors or antenatal care among comparison groups have not identified any factor (or combination) that could explain the improved neonatal survival (Paneth et al., 1982a, 1982b; Paneth, Kiely, and Susser, 1984; Horwood et al., 1982).

Should we insist on a randomized trial? Randomized controlled trials of NIC programs have been recommended in the literature (Sinclair et al., 1981; Sinclair, 1982), but at present are judged unlikely to be carried out for several reasons (Budetti et al., 1981). First, because of the widespread dissemination of NIC, it may be difficult to find a setting where an uncontaminated control group could be maintained. Second, such studies would be very expensive in time and money, given that a reasonably large group of infants would need to be followed for some time. Third, it is not now clear to what extent such trials would influence the dissemination of this technology in developed countries. The data currently available suggest that the question is not whether or not NIC works but how to make it work best by improving the organization and availability of NIC programs.

The third guideline addresses the intervention and whether bias associated with its application was avoided. NIC presents special problems as an intervention. Unlike a drug or many diagnostic tests that are relatively easy to describe in precise, reproducible terms, a NIC program is a multidimensional intervention "package" comprising many procedures and services. Therefore, the actual intensity and sophistication of care available can be highly variable among different centres; it is well known that a centre claiming to offer level 2 care may in fact offer several components or services consistent with level 3 care (Budetti et al., 1981). Hence, distinguishing levels of care (or equating them) among centres may be difficult.

Bias associated with the intervention concerns lack of blindness, co-intervention, contamination, and patient compliance. Ideally, in any evaluation those responsible for administering care should be unaware of which intervention the patient is receiving. Clearly, NIC cannot be administered blind. Contamination of potential comparison groups is another potential problem. Because of the widespread implementation of NIC in developed countries, many authors have stated that it is becoming increasingly difficult to find a population of infants who have not been exposed to some form of NIC (Budetti et al., 1981; Sinclair et al., 1981; Paneth et al., 1982a).

Prenatal and obstetric care also contribute to neonatal outcomes and as such are potential co-interventions. Advances such as inhibition of premature labour resulting in prolonged gestation and greater birth weights and the antenatal administration of steroids to prevent RDS are two factors that could result in decreased newborn risk for mortality and morbidity. One study is available that looked at these factors; neither were found to exert an important influence on mortality differences in infants weighing less than 1,500 grams at birth (Horwood et al., 1982).

Finally, patient compliance can bias estimates of how well an intervention works. In the case of NIC in the Canadian setting it is reasonable to assume that the majority of parents comply with NIC when their physician recommends it for the care of their infant (Sinclair, 1984).

The fourth efficacy guideline addresses the measurement of all relevant outcomes. Both mortality and morbidity (including quality of life) are important in the evaluation of NIC outcomes. Measurement of outcomes should include not only the neonatal period and status at hospital discharge but long-term follow-up to assess subsequent mortality and the development and functional status of survivors.

A key issue in mortality is whether NIC has actually prevented deaths that would have occurred in its absence or simply prolonged survival of sick infants past the neonatal period with little or no effect on mortality overall. Several reports have examined this question (Paneth et al., 1982a; Budetti et al., 1981;

Sinclair, 1982; Horwood et al., 1982) where mortality was assessed during the neonatal period, at the time of hospital discharge, and at later follow-up. The findings are consistent with an overall reduction in mortality, although a slight increase in post-neonatal morbidity has been observed (Horwood et al., 1982; Buehler, Hogue, and Zaro, 1985).

Morbidity assessments including developmental status, functional disability, and quality of life (for both parents and infants) over the long term are crucial in documenting comprehensively the outcome of infants who survive. Several follow-up studies are now available that report on these various outcomes (Boyle et al., 1983; Shapiro et al., 1983).

The McMaster before-after study by Boyle and colleagues (1983) conducted extensive assessments of outcomes among survivors including parental reports of major sequelae and utility measurements of the value associated with the health states (or outcomes) of NIC survivors by a random sample of parents. (These utility measures, used to conduct a cost-utility analysis, are described in detail in the second section of this chapter. In addition, forecasts of future disability and level of functioning were made by developmental pediatric experts (Boyle et al., 1983).

The accuracy and reproducibility of data obtained through interviews or questionnaires requiring parental recall is a concern. Horwood and colleagues (1982) found satisfactory agreement between questionnaire responses and interviews held with the parents six months later. Agreement between parents and physicians on the presence or absence of major sequelae were also found to be highly satisfactory (the chance-corrected kappa [k] statistic for agreement between parents and physicians was 0.72 or 72 per cent).

One problem that remains unsolved is whether the current status of NIC survivors can be used to predict later disability and level of functioning. The study by Boyle and colleagues (1983) forecast outcomes to 15 years of age and to predicted age at death; clearly the validity or accuracy of these forecasts cannot be confirmed without long-term follow-up of the children involved.

Another key outcome of NIC is the trade-off between decreasing mortality and increasing morbidity and the overall burden of illness among survivors of NIC. Is the increase in survival associated with NIC accompanied by an increase in the rate of serious handicap? Evidence on this question comes from long-term follow-up studies of survivors. Most reports have shown that while the proportion of survivors with serious handicap has remained constant or decreased, the absolute number of both normal and impaired survivors has increased (Budetti et al., 1981; McCormick, Shapiro, and Starfield, 1985). The study by Shapiro and colleagues (1983) (which compared a random sample of infants in 1976 with one of infants born in 1978-1979, after a period of rapid growth and promotion of regionalized perinatal care) showed a decrease in developmental delay at one year of age but no substantive change in the moderate or severe categories. The price society currently pays for the increased absolute number of survivors is probably a slightly increased absolute number of handicapped survivors (Budetti et al., 1981). Only long-term follow-up will confirm whether any as yet unknown behavioural or learning difficulties will occur to any significant extent among NIC survivors.

Evidence on the potential toxic effects of NIC has been available for sometime. Retrolental fibroplasia (RLF), or blindness of prematurity, is one of the most well-known examples (Kinsey, 1956; Sinclair et al., 1981). Although it can be partly controlled by careful monitoring of oxygen administration, recent work has suggested that the occurrence of blindness in VLBW infants is becoming epidemic, similar to the time the problem was first identified (Phelps, 1981).

Masking of investigators to the treatment received is essential in the accurate assessment of outcomes subject to measurement bias. Although mortality, the major end point used in evaluating NIC, will not be influenced by lack of masking, assessment of the cause of death, level of functioning, and quality of life could be influenced by knowledge of whether subjects received NIC. Parents of infants not receiving NIC, for example, might exaggerate the severity of their child's handicap or health care needs if they felt that the child would have had a better outcome had he or she received NIC. Or, they might minimize the severity of handicaps. Similarly, interviewers may interpret parental responses and infant developmental assessments differently depending on knowledge of the birth weight of the child or whether he or she received NIC.

Although true masking is difficult because the children may possess obvious characteristics of prematurity or their age may indicate whether they were born in the NIC era, two studies are available that attempted either to minimize the effects of lack of blindness or to measure its effect. The interviewers used by Shapiro and colleagues (1983) were unaware of the birth weight of the infants they assessed, although the authors acknowledge that it may have become obvious that the child was premature. Horwood and colleagues (1982) compared the parents' reports of the presence or absence of major sequelae with physician assessments and found a high level of agreement.

The fifth efficacy guideline focuses on applicability. For NIC, the applicability of evidence depends upon provision of important details about the setting, the patients studied, and the intervention applied. Most of the studies of NIC provide detailed information about the mothers (demographic, socioeconomic, prenatal, and obstetric variables) and infants (clinical status at birth). However, documentation of referral filter biases that may be operating is usually less clear or not provided. A study in New York comparing outcomes according to level of NIC, for example, found mortality in level 3 hospitals was much lower than in level 1 and level 2 hospitals. The authors addressed the issue that mothers may be differentially referred according to perceived risk and level of care available. They could not find any evidence that this was occurring but could not conclude definitively that such biases did not influence the differences in mortality rates observed (Paneth et al., 1982a).

As already discussed, NIC can vary considerably from centre to centre despite the existence of guidelines on levels of care. Therefore, in applying the results of studies it may be more helpful if details of the specific procedures, facilities, and staff available are documented (rather than the use of various labels), allowing for direct comparison by the readers to their own setting.

Finally, factors relating to ability to pay may influence the applicability of the results in certain settings.

The last efficacy guideline asks whether the data were handled appropriately. In the hospital-based randomized trial, all infants entered into the study are reported in the results. Completeness of follow-up in the McMaster study was adequate. The recent morbidity follow-up study by Shapiro and colleagues (1983), however, included only 60 per cent of eligible infants in the two study periods, a rather disturbing rate given that the outcomes of the 40 per cent not followed up could have a profound effect on the study's conclusions. Only limited reassurance is provided by the authors finding that high-response areas did not differ from low-response areas on characteristics found on the birth certificate.

A key issue in the analysis of neonatal and infant mortality is stratification by birth weight. As is recognized in virtually all studies, if the analysis is not conducted within birth weight strata, changes in the birth weight distribution may confound comparisons.

Finally, the clinical importance of the results must be considered to determine whether the differences observed would lead one to recommend NIC for the care of specific infants. In the Kitchen Trial (Kitchen and Campbell, 1971), 18 per cent of infants receiving routine care died in the neonatal period compared with 35 per cent of those receiving routine care. This constitutes a 49 per cent decrease in mortality, clearly of clinical importance. The failure to obtain statistical significance in this study may have been due to insufficient sample size or power.

The results of the Boyle study showed an overall decrease in neonatal mortality for VLBW infants from 58.7 per cent to 39.2 per cent – a 19.5 per cent absolute decrease or a 33 per cent risk reduction in mortality. Examination of the results by birth weight group showed that the mortality reduction was greater in the 1,000-1,499 than in the 500-999 gram group (1,000-1,499 gram group, 36.6 per cent to 21.6 per cent or 41 per cent risk reduction; 500-999 gram group, 88.1 per cent to 69.4 per cent or 21 per cent risk reduction). The clinical importance of this difference depends upon whether it would lead one to recommend NIC for all moderate to high-risk infants or a subset. We would suggest that the magnitude of the difference is such that the dissemination of NIC is justified.

Turning to the data available on the individual components that determine community effectiveness (efficacy, diagnostic accuracy, patient compliance, provider compliance, and coverage), information on coverage is available from two studies. Boyle and colleagues (1983) found that 68 per cent of all liveborn VLBW infants received NIC during the "after" study period when it was available (Horwood et al., 1982). Although it is not known what proportion of the remaining 32 per cent needed NIC and did not get it, in the Canadian setting it is likely that in the majority of cases the infant died in the first few hours after birth before transfer to a level 3 centre could be arranged (Sinclair, 1984). Paneth and colleagues (1982a; Paneth, Kiely, and Susser, 1984) compared mortality between three New York City hospitals offering different levels of newborn care and found that mortality rates were significantly lower in the level 3 facility than in either the level 1 or 2 hospitals. The authors could not identify any confounding factor (such as unequal distribution of social, demographic, or biologic risk factors across the hospitals) that could explain the difference in mortality. They concluded that efforts are needed to optimize the effectiveness of NIC, such as appropriate transfer of infants to level 3 facilities and ensuring that as many low birth weight infants as possible are delivered in hospitals with level 3 facilities. Further investigation of the individual components of community effectiveness will help to identify factors below the minimum level judged adequate for the purpose of undertaking corrective action.

Finally, the importance of continued, serial evaluation of a health technology is underlined by the recent publication of Saigal and colleagues (1984). They report follow-up data on a more recent cohort of infants weighing 500-999 grams at birth – all those liveborn to residents of the McMaster Health Region between 1977 and 1980. Their data show that survival among these infants has doubled, compared with their earlier results for infants born between 1973 and 1977, from 24 per cent to 49 per cent. No important change in morbidity among survivors has accompanied this decrease in mortality. Further analysis of their findings suggests that this improved picture is particularly important for infants weighing 700 grams or more at birth (less than 700 grams, 18 per cent survival; more than 700 grams, 62 per cent survival). Their results are consistent with Paneth and colleagues (1982a; Paneth, Kiely, and Susser, 1984), in that overall mortality among infants (500 to 999 grams) born at the level 3 centre was lower than that for infants born at community hospitals in the McMaster Health Region.

Economic Evaluation

The first section of this chapter presented evidence on effectiveness but did not discuss estimates of the economic and psychic costs and benefits of NIC. What are the costs of increasing the survival rate? What is the quality of life of the survivors? How do the costs of treating low birth weight neonates compare with the costs of reducing the incidence of low birth weight among neonates – the costs of prevention? These issues will now be explored.

A number of studies have documented the high economic costs of treatment in NIC units. Although some of the studies omit physician fees and some measure only a portion of the hospital costs, they are almost unanimous in pointing to the high costs of NIC – hospital costs of the same order of magnitude as coronary artery bypass surgery (Budetti et al., 1981; see also Kaufman and Shepard, 1982; McCarthy, Honeyfield, and Butterfield, 1979; Phibbs, Williams, and Phibbs, 1981; Pomerance et al., 1978; Walker et al., 1984; and Yu and Bajuk, 1981).

Several patterns emerge from the cost studies (efficiency guideline 7). First, cost per day of stay in a NIC is initially high and decreases over time (Kaufman and Shepard, 1982; Budetti et al., 1981). Second, the direct costs of treating the infant and operating the NIC unit itself account for roughly three fourths of the total cost of providing care (Kaufman and Shepard, 1982). Third, in general, costs increased rapidly with decreasing birth weight for infants below 2,000 grams (Phibbs, Williams, and Phibbs, 1981). Finally, other factors, such as whether the newborn underwent a significant degree of surgery and whether the infant received more than twenty-four hours of assisted ventilation, were also significant in explaining the substantial variation in costs per patient (Phibbs, Williams, and Phibbs, 1981).

The studies discussed above provide benchmarks of the costs and in some cases mortality and/or morbidity reductions associated with NIC. They do not, however, satisfy efficiency guideline 2 by presenting a rigorous economic evaluation from a societal point of view in which the costs, benefits, and quality of life effects of the NIC are compared. Fortunately, the paper by Boyle et al. (1983) provides the kind of economic evaluation needed to assess the effectiveness and efficiency of NIC. The analysis was conducted for very-low birth weight infants (less than 1,500 grams). The point of view was that of society. As discussed in the first section, the approach was to compare the costs and outcomes for the target population (infants born to residents in a defined region) before and after the introduction of regional NIC facilities. This illustrates efficiency guideline number 3, consideration of a relevant range of alternatives. Techniques of cost-effectiveness, cost-utility, and cost-benefit analyses were applied, all of which are appropriate forms of economic evaluation for the questions asked (efficiency guideline 4). To conduct the cost-utility analysis, measures of health states were developed for survivors according to their physical function, role function, social and emotional function, and health problems. A random sample of parents in the region was drawn and asked to attach relative values to each of the health states in the classification. In this utility measurement a value of "1" was assigned to a healthy state and a value of "0" to death as an extreme state of health. Parents rated some of the chronic dysfunctional states in children as being worse than death, however, and the scale had to be extended from 1 to -0.39. The utilities elicited from the sample of parents were used in the cost-utility analysis to evaluate the quality of life of the outcomes.

The costs of NIC care and the care provided before NIC was available were comprehensively identified and carefully computed satisfying efficiency guidelines 5 and 6. The costs include the direct costs of the NIC unit itself, the costs of services provided to the unit by other hospital departments (such as radiology), and overhead items (such as buildings, equipment, administration, and housekeeping). Physician charges, the costs of convalescent care in community hospitals, and the cost of ambulance transport to and from the regional NIC facilities were also included. In short, the direct economic costs of providing neonatal intensive care and the pre-NIC treatments were accurately estimated.[1]

In addition to the costs for NIC or pre-NIC treatment, follow-up costs were also estimated. They included health care costs (such as hospital, medical, dental, and drugs) and other costs (such as institutional costs, special education). A measure of the opportunity costs of parental time for the care of handicapped but non-institutionalized patients was included. Costs incurred in the future were translated into their present values (efficiency guideline 8) through the use of a 5 per cent discount rate (with sensitivity tests of 0 through 10 per cent, efficiency guideline 9).

The cost of care estimates provided by Boyle and colleagues (1983) have the advantage of being more comprehensive, careful, and accurate than those presented in the previous studies already discussed. The estimates accurately measure the social opportunity cost of providing care. They do, however, omit several other costs that the analyst may wish to consider. First, parents of patients in a NIC unit may experience some loss of income (or more appropriately, society may experience some loss of output) because of time taken off work to attend to the care of their infants. Similarly, to the extent that the level of home care provided to infants after discharge exceeds the care requirements of normal infants (other than those classified as handicapped) and to the extent that the "excess" care causes parents to reduce their labour force participation (or forgo some leisure), there may be some additional costs not captured in the estimates. Second, there are psychic costs for parents of infants being treated in the NIC unit. The Boyle study does an excellent job of assessing the quality of life of the health outcomes. It does not, however, provide evidence on the quality of life (for the parents or the patients) during the process of treatment by which those outcomes are determined.[2]

The effects of NIC were also carefully considered. The actual survival rates experienced by the 1964-1969 and 1973-1977 groups were compared. Among infants weighing 1,000 to 1,499 grams at birth, 77.2 per cent survived to hospital discharge in the later era versus 62.4 per cent in the earlier period. For those weighing 500 to 999 grams, the survival rate increased from 10.6 per cent to 22.4 per cent. Although the reduction in mortality rates experienced for the 1,000-1,499 gram group is roughly comparable to the results reported for many other centres (Budetti et al., 1981), higher rates of mortality reduction have been reported in several other studies for the 500-999 gram group (Budetti et al., 1981; Yu and Bajuk, 1981; and Kaufman and Shepard, 1982). The results of these later studies are not directly comparable, however, because they are hospital based rather than area based. They include outcomes only for infants who reached hospital and as a result may present a biased estimate. Knowledge of whether or not the "effect" data used in an efficiency estimate was derived from a hospital-based or area-based study is crucial to examining the potential for bias. When compared with other area-based studies, the decline in mortality was larger in the Hamilton-Wentworth area (see Horwood et al., 1982). Because the economic evaluation of NIC performed by Boyle and colleagues (1983) has the advantage of being based on the actual decline in neonatal mortality experienced in the geographic area, there is every reason to believe that the benefits of NIC are accurately measured.

The effects on mortality were assessed by comparing the experience before and after the introduction of the regional NIC facilities. Increased survival was then measured in the number of life-years gained. Because significant chronic dysfunction or morbidity is associated with some of the survival outcomes, the utility values were applied to weigh the life-years gained and arrive at quality-adjusted life-years (QALYs) gained. Finally, the economic benefits of the intervention were estimated by comparing the projected lifetime earnings of survivors of NIC against the experience of low birth weight neonates before the introduction of NIC. Lifetime earnings were estimated on the basis of age-specific and sex-specific earnings derived from census data and discounted to their present values.

The procedures were sound and reasonable but might tend to underestimate lifetime earnings for two reasons. First, if female labour force participation rates increase over the lifetime of the surviving patients, the use of age- and sex-specific earnings from an earlier period during which female labour force participation rates were lower might bias downwards the lifetime earnings of females. Second, in a related manner, if the earnings of some groups in society are artificially low because those groups face discrimination in the labour market, their actual earnings may understate their potential earnings, again imparting a downward bias in the estimates. This potential bias may be especially relevant in evaluating neonatal care in that the mentally and physically handicapped may in fact face such discrimination in the labour market. While sensitivity tests do not appear to have specifically addressed these problems, the manner in which the authors compare their costs and benefits makes their results insensitive to these potential sources of bias.

The reason that the overall results are insensitive to these problems is that the authors compare the incremental costs and benefits of NIC. They compute the additional costs incurred in treating patients after the introduction of NIC and compare them with the incremental benefits in additional life-years gained, quality-adjusted life-years gained, and additional lifetime earnings. Thus, if lifetime earnings have been underestimated, they were underestimated for the "control" and "treated" groups (the before and after groups). The true difference between the two is likely to have been accurately measured.

The study by Boyle and colleagues (1983) is appropriately designed to answer the question of how worthwhile NIC is relative to the traditional mode of treatment.[3] The benefits of the gains in knowledge and skills from providing treatment are difficult to measure and therefore were omitted.

Although the study by Boyle is the best and most comprehensive available, it does not consider the macro or systems issues of the effect of the availability of NIC (efficiency guideline 1). It carefully examines the costs and effects of the intervention itself but provides no evidence on other impacts, such as the effect on the willingness of high-risk parents to have children. These broader issues do not appear to have been addressed in the literature on NIC.

The conduct of the analysis and presentation of results illustrate guidelines 8, 9, and 10—allowance for differential timing of costs and benefits, allowance for uncertainty in costs and benefits, and appropriate presentation of results.

Tables 10.1, 10.2, 10.3, and 10.4, taken from the study by Boyle and colleagues (1983), present the results of their cost-effectiveness, cost-utility, and cost-benefit analyses. The introduction of NIC appears to have increased survival substantially as reflected in the increases in survival rates and life-years per live birth (see Table 10.1). These gains are consistently attenuated when one adjusts for quality of life, reflecting long-term disability of survivors. This is especially true for infants weighing less than 1,000 grams at birth.

The increases in life-years and QALYs are obtained, however, at a considerable increase in the costs of both hospital treatment in the NIC unit and subsequent health care costs. Table 10.4 presents the net economic cost estimates (discounted at 5 per cent). For the 1,000-1,499 gram group, the increased costs of treatment are very nearly offset by the increased lifetime

TABLE 10.1

Health Outcomes for Very Low Birth Weight Infants Born Before and After the Introduction of Neonatal Intensive Care, by Birth Weight Class (Undiscounted)

	Birth Weight 1,000-1,499 grams		Birth Weight 500-999 grams	
	Before Intensive Care[a]	Intensive Care[a]	Before Intensive Care[a]	Intensive Care[a]
Period	(n=213)	(n=167)	(n=160)	(n=98)
To hospital discharge				
Survivors (%)	62.4	77.2	10.6	22.4
To age 15 (projected)				
Life-years/live birth	9.0	11.1	1.46	3.37
QALYs[b]live birth	6.4	8.1	1.22	1.80
To death (projected)				
Life-years/live birth	38.8	47.7	6.6	13.0
QALYs[b]live birth	27.4	36.0	5.5	9.1

a. The period studied before the introduction of neonatal intensive care is 1964-1969; the period after, 1973-1977.

b. QALY denotes quality-adjusted life years.

Source: Boyle et al. (1983) p. 1332.
Reprinted by permission of the *New England Journal of Medicine*.

TABLE 10.2

Economic Outcomes by Birth Weight Class (Undiscounted; 1978 Canadian Dollar)

Cost/Live Birth	Birth Weight 1,000-1,499 grams		Birth Weight 500-999 grams	
	Before Intensive Care[a]	Intensive Care[a]	Before Intensive Care[a]	Intensive Care[a]
To hospital discharge				
Health care	5,400	14,200	1,500	13,600
To age 15 (projected)				
Health care	8,100	18,700	1,800	18,000
Other	4,000	2,000	200	1,900
Total	12,100	20,700	2,000	19,900
To death (projected)				
Health care	45,400	61,500	9,500	28,600
Other	47,100	38,600	1,500	15,000
Total	92,500	100,100	11,000	43,600
Earnings/Live birth				
To death (projected)	122,200	154,500	19,200	48,100

a. The period studied before the introduction of neonatal intensive care is 1964-1969; the period after, 1973-1977.

Source: Boyle et al. (1983) p. 1332.
Reprinted by permission of the *New England Journal of Medicine*.

TABLE 10.3

Measures of Economic Evaluation of Neonatal Intensive Care, by Birth Weight Class (Undiscounted; 1978 Canadian Dollar)

	Birth Weight Class	
Period	1,000-1,499 grams	500-999 grams
To hospital discharge		
Cost/additional survivor at hospital discharge	59,500	102,500
To age 15 (projected)		
Cost/life-year gained	4,100	9,400
Cost/QALY gained[a]	5,100	30,900
To death (projected)		
Cost/life-year gained	900	5,100
Cost/QALY gained[a]	900	9,000
Net economic benefit (loss)/live birth	24,700	(3,700)
Net economic cost/life-year gained	NA[b]	600
Net economic cost/QALY gained[a]	NA[b]	1,000

a. QALY denotes quality-adjusted life-years

b. NA denotes not applicable.

Source: Boyle et al. (1983), p. 1333. Reprinted by permission of the *New England Journal of Medicine*.

TABLE 10.4

Measures of Economic Evaluation of Neonatal Intensive Care, by Birth Weight Class
(5 per cent discount rate; 1978 Canadian Dollar)

Period	Birth Weight Class	
	1,000-1,499 grams	500-999 grams
To hospital discharge[a]		
Cost/additional survivor at hospital discharge	59,500	102,500
To age 15 (projected)		
Cost/life-year gained	6,100	12,200
Cost/QALY gained[b]	7,700	40,100
To death (projected)		
Cost/life-year gained	2,900	9,300
Cost/QALY gained[b]	3,200	22,400
Net economic benefit (loss)/live birth	(2,600)	(16,100)
Net economic cost/life-year gained	900	7,300
Net economic cost/QALY gained[b]	1,000	17,500

a. All costs and effects occurred in year one.

b. QALY denotes quality-adjusted life-years

Source: Boyle et al. (1983), p. 1333.
Reprinted by permission of the *New England Journal of Medicine*.

earnings. In fact, the sensitivity tests performed by the authors indicate that at a discount rate of 3 1/2 per cent (or lower), the net economic costs are negative, that is, the increased earnings more than compensate for the incremental treatment costs. Given the assumption that in the absence of the NIC unit the health care sector (and society) would in fact choose to offer treatment to very low birth weight neonates, this is a powerful argument in favour of the use of NIC unit technology for that group of patients. (In the period before the introduction of the NIC unit, VLBW infants did in fact receive care in the regular nursery.)

For the 500-999 gram group, the results are not as favourable. The large increase in cost represented by the introduction of the NIC unit results in some decreases in mortality and increases in life years, but much smaller gains in QALYs. The net economic costs are clearly positive; it costs more to care for these patients than can be repaid by their lifetime earnings. More important, some of these survivors will have very low quality of life.

A number of issues are raised by the results. First, as the authors suggest, in any situation in which rationing of the use of current facilities is necessary, all three economic criteria (cost-effectiveness, cost-utility, cost-benefit) indicate that 1,000-1,499 gram patients might be favoured over the lower birth weight group. The conclusion follows from comparisons of the cost per life-year gained, cost per QALY, and net economic benefit per live birth.

Second, the role of the lower survival rate experienced by the 500-999 gram group in accounting for the high cost per life-year gained needs to be assessed. As explained above, however, the survival rate experienced in the Hamilton-Wentworth region was, at the time, on the high end of the range. Thus it seems reasonable to conclude that if there is any bias in the Boyle results it is toward overestimating the effectiveness of NIC for the 500-999 gram group. One is forced to conclude, as did Budetti and colleagues (1981) that the cost per life-year gained or quality adjusted life-year gained is indeed truly much higher for the 500-999 gram group.

The results from Saigal and colleagues (1984) for the Hamilton-Wentworth region demonstrated an increase in survival for the 500-999 gram group. This new information cannot unambiguously be used, however, to argue that the cost per life-year or quality-adjusted life-year gained is less than previously thought, as costs may have changed, too, and we lack up-to-date cost estimates.

Third, given that conclusion, what are the implications for future research and resource allocation? One strategy would be to compare the cost per QALY and net economic benefits of NIC for 500-999 gram neonates to similar measures derived from an evaluation of prenatal care programs designed to reduce the incidence of very low birth weights among neonates. Improved nutrition and prenatal medical care, improved labour and delivery for high-risk individuals, and effective anti-smoking measures might be reasonable strategies. Unfortunately, there do not appear to be any rigorous evaluations of these interventions with which to compare the NIC results.[4]

Another strategy is to design and test intermediate treatment regimens for the 500-999 gram group in an effort to find lower-cost ways of achieving the same outcomes.[5] A related strategy is to investigate ways to predict accurately which neonate patients from this group are most likely to survive and, in particular, experience higher QALYs?.

Still another avenue of inquiry might be the quality of care received by these very small infants. It may be possible to identify deficits in care, such as pre-delivery transfer of infants likely to be born weighing 1,000 grams or less, which if improved could result in improved mortality and morbidity outcomes. The McMaster study found that 32 per cent of all VLBW infants did not receive NIC. If these were infants weighing 1,000 grams or less, increases in coverage might significantly improve outcomes in this weight group. Issues of the optimal degree of utilization, alternative forms of organization of NIC, and alternative regional facility plans and systems of delivery are still unresolved.[6]

Finally, given that technologies evolve and generally improve over time, some form of serial technology assessment and updating of earlier evaluations is needed. The result of the study by Saigal and colleagues (1984) reviewed above are useful in this regard. In the period since 1973-1978, survival rates among VLBW infants including the under 1,000 gram group have improved; it is likely that costs have risen as well. Thus the results of the study by Boyle and colleagues (1983) (or any study) should be adjusted to reflect new evidence in a timely manner for use in policy making.

A major question raised by the findings for the 500-999 gram group (and findings in general) is how much society is willing to pay for a QALY. The Boyle study shows that the net economic cost of NIC for the 500-999 gram group is positive. Many other health programs probably have similar outcomes. Although the stock of high-quality cost-utility analyses is not sufficient to make a comprehensive comparison between NIC and other programs, such as dialysis, kidney transplantation, geriatric programs, and treatment of childhood cancers, it is not clear that society is spending more per QALY in neonatology than it is in other areas.[7] The important point is that we need the information and analysis base upon which to make consistent choices among programs so that the net economic cost per QALY may be roughly equalized or deviations from that standard rigorously defended.

Given the difficult ethical issues posed by clinical and economic evaluation of NIC and the complicating quality-of-life issues that are intrinsically involved, another implication is that a mechanism for social choice is needed to provide guidance to physicians. The question is how much is society willing to give up to improve health outcomes? Neither physicians nor health systems analysts have any special expertise in fundamental social value judgements.

Conclusion

We believe the following conclusions are warranted based on the foregoing review of NIC programs within our guidelines. Although NIC has not (with one exception) been subjected to evaluation through the use of randomized clinical trials, a number of the manoeuvers that constitute NIC have. These manoeuvers appear to be effective, and there is widespread evidence pointing to the effectiveness of NIC programs in accounting for much of the recent decline in birth-weight-specific neonatal mortality rates (Budetti et al., 1981; Horwood et al., 1982). It is reasonable to conclude that NIC is an effective intervention for the care of low birth weight neonates, especially for the 1,000-1,499 gram group. For the 500-999 gram group, increased survival rates are less spectacular, and NIC is less effective in terms of the resulting quality of life for survivors. The costs are also much higher. The provision of regional neonatal intensive care services for the former group would appear to be socially desirable by any of the criteria used. For the latter group the results are less clear, and several options need to be explored further. However, the new evidence on improved outcomes in this lower weight group, particularly for infants weighing more than 700 grams at birth, is encouraging (Saigal et al., 1984). We believe that these general conclusions hold for almost any developed country; absolute magnitudes may vary depending upon local circumstances.

Although the guidelines are not intended to provide answers for the planning of health services, the conclusions should contribute significantly to decisions about the utilization of existing neonatal intensive care programs as well as the implementation of NIC programs in areas not currently served.

We recommend that the available evidence from other health care technologies should be synthesized and analysed against methodological guidelines in the manner illustrated here.

Notes

1. A very minor caveat is that the interest rate used to estimate the cost of equipment usage was 9.46 per cent (the bond rate then current on Ontario Hydro long-term bonds), instead of the 5 per cent rate used in computing the present values of the projected health care costs and benefits. Given the fairly low share of equipment costs in total costs and the fairly minor effect that substituting 5 per cent for 9.64 per cent would have on the equipment cost estimates, the likely effect on the overall results is minimal and is in some sense handled by the sensitivity tests on the discount rate that were performed.

2. The quality of life measures for the health outcomes were based on the evaluations provided by a representative sample of parents in the Hamilton-Wentworth region. They do not, however, necessarily accurately reflect the assessments of siblings of survivors of the NIC unit of the former patients themselves.

3. The study by Boyle and colleagues (1983) is not specifically focused on answering the much less relevant question of the costs and benefits of the NIC relative to a non-treatment option in which no costs are incurred.

4. There is at least one on-going randomized controlled trial addressing the evaluation of prenatal interventions in Canada and another in Western Australia.

5. Bose, Lapine, and Fung (1985) present some suggestive results on the potential cost savings of transferring convalescing infants from level 3 units to community hospitals with level 1 or 2 facilities. Such a strategy is, however, of more relevance in reducing the costs for the 1,000-1,500 gram group than for the under 1,000 gram group.

6. Some of these issues are discussed in Siegel et al., 1977. See also Paneth et al., 1982a and Paneth, Kiely, and Susser, 1984.

7. Evidence on the cost per QALY gained for a variety of health care interventions can be found in Torrance and Zipursky (1984) and Torrance (1986). For arguments on why society appears to be willing to spend more on treatment than prevention and more on identified patients ex post than on "statistically expected" patient ex ante, see Weinstein, Shepard, and Pliskin, 1980.

References

Bloom, B.S. 1984. "Changing Infant Mortality: The Need to Spend More While Getting Less." *Pediatrics* 73(6): 862-866.

Bose, C.L.; LaPine, T.R.; Fung, A.L. 1985. "Neonatal Back-transport: Cost Effectiveness." *Medical Care* 23(1): 14-19.

Boyle, M.H.; Torrance, G.W.; Sinclair, J.C.; Horwood, S.P. 1983. "Economic Evaluation of Neonatal Intensive Care of Very-Low-Birth-Weight Infants." *New England Journal of Medicine* 308: 1330-1337.

Budetti, P.; Barrand, N.; McManus, P.; Heinen, L.A. 1981. "The Implications of Cost-Effectiveness Analysis of Medical Technology. Background Paper #2: Effectiveness of Neonatal Intensive Care." Washington, D.C.: Office of Technology Assessment.

Buehler, J.W.; Hogue, C.J.R.; Zaro, S.M. 1985. "Postponing or Preventing Deaths? Trends in Infant Survival, Georgia, 1974 through 1981." *Journal of the American Medical Association* 253(24): 3564-3567.

Committee on Perinatal Health. 1976. "Toward Improving the Outcome of Pregnancy: Recommendations for the Regional Development of Maternal and Perinatal Health Services." White Plains, N.Y.: The National Foundation – March of Dimes.

Horwood, S.D.; Boyle, M.H.; Torrance, G.W.; Sinclair, J.C. 1982. "Mortality and Morbidity of 500 to 1,499 Gram Birth Weight Infants Liveborn to Residents of a Defined Geographic Region before and after Neonatal Intensive Care." *Pediatrics* 69: 613-620.

Kaufman, S.L., and Shepard, D.S. 1982. "Costs of Neonatal Intensive Care by Day of Stay." *Inquiry* 19: 167-178.

Kinsey, V.E. 1956. "Retrolental Fibroplasia: Co-operative Study and the Use of Oxygen." *Archives of Opthalmology* 56: 481-543.

Kitchen, W., and Campbell, D. 1971. "Controlled Trial of Intensive Care for Very Low Birthweight Infants." *Pediatrics* 48: 711-714.

Kleinman, J.C.; Kovar, M.G.; Feldman, J.J.; Young, C.A. 1978. "A Comparison of 1960 and 1973-1974 Early Neonatal Mortality in Selected States." *American Journal of Epidemiology* 108: 454-469.

Lee, K.W., Paneth, N.; Gertner, L.M.; et al. 1980. "Neonatal Mortality: An Analysis of the Recent Improvement in the United States." *American Journal of Public Health* 70: 15-21.

McCarthy, J.T.; Honeyfield, P.R.; Butterfield, L.J. 1979. "Who Pays the Bill for Neonatal Intensive Care?" *Pediatrics* 95: 755-761.

McCormick, M.C.; Shapiro, S.; Starfield, B.H. 1985. "The Regionalization of Perinatal Services: Summary of the Evaluation of a Neonatal Demonstration Program." *Journal of the American Medical Association* 253(6): 799-804.

Paneth, N.; Kiely, J.L.; Susser, M. 1984. "Age of Death Used to Assess the Effect of Interhospital Transfer of Newborns." *Pediatrics* 73(6): 854-861.

Paneth, N.; Kiely, J.L.; Wallenstein, S.; et al. 1982a. "Newborn Intensive Care and Neonatal Mortality in Low Birth Weight Infants." *New England Journal of Medicine 307: 149-155.*

Paneth, N.; Wallenstein, S.; Kiely, J.L.; Susser, M. 1982b. "Social Class Indicators and Mortality in Low Birth Weight Infants." *American Journal of Epidemiology* 116: 364-375.

Phelps, O.L. 1981. "Retinopathy of Vision Loss in the United States." *Pediatrics* 67: 924-926.

Phibbs, C.S.; Williams, R.L.; Phibbs, R.H. 1981. "Newborn Risk Factors and Costs of Neonatal Intensive Care." *Pediatrics* 68: 313-321.

Pomerance, J.J.; Ukrainski, C.T.; Ukra, T.; et al. 1978. "Cost of Living for Infants Weighing 1,000 Grams or Less at Birth." *Pediatrics* 61: 908-910.

Saigal, S.; Rosenbaum, P.; Stoskopf, B.; Sinclair, J.C. 1984. "The Outcome of Infants 500-1,000 Grams Birth Weight Delivered to Residents of the McMaster Health Region." *Journal of Pediatrics* 105: 969-976.

Shapiro, S.; McCormick, M.C.; Starfield, B.H.; Crawley, B. 1983. "Changes in Infant Mortality Associated with Decreases in Neonatal Mortality." *Pediatrics* 72: 408-415.

Siegel, E.; Gillings, D.; Guild, P.; Nugent, R. 1977. "Planning and Evaluation of Regionalized Perinatal Care: A Rural Example." *Seminars in Perintology* 1: 283-301.

Sinclair, J.C. 1982; "Effectiveness of Intensive Care of Very Low Birth Weight Infants." In *Clinical and Economic Evaluation of Perinatal Programs*, Mead Johnson Symposium on Prenatal and Developmental Medicine, No. 20, June.

Sinclair, J.C. 1984. Personal Communication.

Sinclair, J.C.; Torrance, G.W.; Boyle, M.H.; et al. 1981. "Evaluation of Neonatal Intensive Care Programs." *New England Journal of Medicine* 305: 489-494.

Torrance, G.W. 1986. "Measurement of Health State Utilities for Economic Appraisal: A Review." *Journal of Health Economics* 5(1): 1-30.

Torrance, G.W., and Zipursky, A. 1984. "Cost Effectiveness of Antepartum Prevention of Rh Immunization." *Clinics in Perintology* 11(2): 267-281.

Walker, D-J.B.; Feldman, A.; Vohr, B.R.; Oh, W. 1984. "Cost-Benefit Analysis of Neonatal Intensive Care for Infants Weighing Less Than 1,000 Grams at Birth." *Pediatrics* 74(1): 20-25.

Weinstein, M.C.; Shepard, D.S.; Pliskin, J.S. 1980. "The Economic Value of Changing Mortality Probabilities: A Decision-Theoretic Approach." *Quarterly Journal of Economics* 94: 373-396.

Williams, R.L., and Chen, P.M. 1982. "Identifying the Sources of the Recent Decline in Perinatal Mortality Rates in California." *New England Journal of Medicine* 306: 207.

Yu, Y.H., and Bajuk, B. 1981. "Medical Expenses of Neonatal Intensive Care for Very Low Birth Weight Infants." *Australian Pediatrics Journal* 17: 183-185.

Chapter 11:

Policy Options for Health Care Technology

Greg Stoddart & David Feeny

Introduction

The preceding chapters have shown that technology assessment is indeed a complex issue. The assembly and due consideration of evidence is both important and possible in the process of technology assessment, but for a variety of reasons, it is often difficult. Unfortunately, the evidence has frequently been ignored or set aside when decision makers have been confronted with other, more pressing problems.

Yet the case for some systematic process of technology assessment is quite strong. From the discussion and literature reviewed in earlier chapters, it is clear that:

1. Technologies are often accepted for general use without evaluation.
2. Technologies are often accepted for use before evaluations are completed, making it extremely difficult to act upon subsequent results when they suggest the technology is not of benefit.
3. Technologies are often over-supplied relative to reasonable estimates of the health needs of the population.
4. Technologies that have been evaluated, and accepted for use on the basis of the evaluation, are often utilized for conditions beyond those covered by the evaluation.
5. Technologies that have been evaluated, accepted for use, and used within the broad conditions for which they have been evaluated are often utilized in the aggregate at frequencies beyond those known to be efficacious and necessary.
6. Even for technologies that avoid the above characteristics, there are situations in which existing utilization of some technologies is cost ineffective in that an equally effective and lower-cost technology exists but is not utilized.

The main theme of this book has been the specification of guidelines for evidence on safety, efficacy, effectiveness, and efficiency. Three important questions remain. First, how is such evidence to be assembled, synthesized, and disseminated so that appropriate policies on technology may be formulated and executed? Second, how can guidelines such as those suggested in earlier chapters be implemented in the diffusion and utilization of health care technology? Third, to what extent does the current system of organizing and financing health care help or hinder these processes?

Appropriate diffusion and appropriate utilization are the two main policy goals. Appropriate diffusion means that new health care technologies are first introduced for research and evaluation and then disseminated for wider application in clinical situations on the basis of the evidence from the earlier rigorous clinical and economic evaluations. Appropriate diffusion would ensure that major new technologies are provided in amounts that correspond to health care needs (constrained by society's ability and willingness to pay) and that such technologies are appropriately located within a jurisdiction.

Appropriate utilization means the application of a technology to cases and conditions where its use is known to be effective. It implies the elimination of the technological component of what Enthoven (1978) has called "flat of the curve" medicine, services that do not contribute to health. Appropriate utilization would ensure that cost-effective technologies replace other technologies. There is a tendency for new technologies that are complementary to existing health care resources to be added on to the system; those that eliminate the need for existing services or reduce health care expenditures (which are also, for the most part, incomes to someone in the health care system) are acted on much less frequently and less completely. Perhaps most important, appropriate utilization implies that information on technologies that are both more costly and more effective (a situation that arises frequently) is made available in a manner that places the difficult choices for decision makers "out in the open," thereby encouraging informed discussion.

Attempts to encourage appropriate diffusion and appropriate utilization will, of course, have to work within a set of constraints. The equitable distribution and provision of services to different geographic and socio-economic groups, for example, have always been important goals of the Canadian health care system. Furthermore, the information required for technology assessment will almost always be incomplete, if for no other reason than that technological change moves more rapidly than the pace of rigorous evaluation.

The timeliness of information is frequently a contentious issue in health care technology assessment. Because evaluation results are often needed quickly, guidelines such as those suggested in Chapters 4 through 8 may be seen as too demanding and unrealistic; however, they represent an important paradigm. Any specific case will require a balancing of the need for evidence against the cost of obtaining it, the possibility of obtaining it in a relevant time period, and the potential social cost (and perhaps even harmful health effects) of not obtaining the necessary evidence. Furthermore, the specific characteristics of different cases will require decision makers to use different blends of evidence and judgement. Nevertheless, there seems little question that more frequent attempts to use guidelines such as those generated in this book will make the decision making process a better informed one.

Perhaps the most serious constraint on any attempt to improve the diffusion and utilization patterns for health care technology is that few, if any, incentives exist in the current structure of the health care system for either appropriate diffusion or appropriate utilization. It seems fair to say that the vested interests of manufacturers, hospital administrators, physicians, and consumers (both individually and collectively as communities) all reinforce tendencies toward technological expansion. Of course, it is irrational to expect these parties to act other than in their own interest; however, as will be discussed below, the current organization and financing of Canadian health care gives none of the actors an opportunity or incentive to align their own goals with a social interest in appropriate diffusion and utilization. In fact, existing institutional arrangements often deter actions that would be in the social interest.

The purpose of this chapter is to identify and to discuss policy options available for accomplishing the goals of appropriate diffusion and utilization of health care technology. The advantages and disadvantages of various policy options are discussed and recommendations are made in order to frame and to stimulate public discussion about technology assessment. While it is tempting to focus on specific technologies and to make specific recommendations in certain cases, technology assessment is broader than any specific technology or any specific technical exercise. It is a process that goes on continuously and is implicit (or should be) in the everyday decisions made by several groups of key actors within the health care system. Therefore, policies designed to create a state of mind or an attitude toward the introduction and use of technology may be more important than higher profile legislative policies, or case-specific policies. Effective change in behaviour is more likely to result from changes in incentives and attitudes than from coercive measures.

Two themes run through the chapter. First, several policy options address the need for technology assessment directly; these can be considered independently of the broader set of issues on the organization, financing, and regulation of the Canadian health care system as a whole. Second, the desire to improve upon existing patterns of diffusion and utilization leads to an examination of these broader issues, this time from the perspective of technology policy rather than that of expenditure control or other questions that have typically motivated examination of the same issues. These two themes will be interwoven in the discussion that follows. The current policy approach in Canada, institutions for the production, synthesis, and dissemination of information, policy options within the existing health care system, and policy options that involve significant changes to the structure of the existing system will be discussed in turn.

The Current Policy Approach to Technology Assessment in Canada: The Information Base

Because there are few formal policy statements or guidelines on technology assessment in Canada, it is somewhat misleading to speak of a policy on technology assessment. It is possible, however, to describe the current approach to technology assessment both with respect to the assembly and synthesis of information on safety, efficacy, effectiveness, and efficiency and with respect to the attempt to ensure appropriate diffusion and utilization.

No co-ordinated national effort and few provincial efforts exist that are aimed at assembly and synthesis of information on a comprehensive or ongoing basis.[1] Studies of varying quality and rigour are undertaken by manufacturers and individual researchers or institutions. Some of these studies are sponsored by the federal government (through research grants issued by agencies such as the Medical Research Council and the National Health Research and Development Program) and the provincial governments (through research councils, direct government grants, or as conditions of approval for new technologies). Others are sponsored by manufacturers, foundations, and individual health care institutions. Although the size of this research effort is an issue of interest, its fragmentation makes it exceedingly difficult to determine a dollar value of the effort. Available data on total health research expenditures in Canada, which include expenditures on technology assessment and on other types of health research, are presented in Table 11.1.

TABLE 11.1

Expenditures on Health Research in Canada 1970-1980

Total private and public sector expenditures on health research as a percentage of total private and public sector expenditures on health	1.09
Public sector expenditures on health research as a percentage of total public sector health expenditures	1.17
Federal government expenditures on health research as a percentage of total federal government expenditures on health	2.15[a]
Provincial government expenditures on health research as a percentage of provincial government expenditures on health	0.46[a]
Private sector expenditures on health research as a percentage of total private expenditures on health	0.82
Total private and public sector expenditures on health research as a percentage of Gross National Product	0.08
Total expenditures on health as a percentage of Gross National Product	7.33

a. Data are for 1970-1978 only.

Source: Based on expenditure data published by Health and Welfare Canada (1981, 1983).

Some attempts have been made to bring together working parties to synthesize existing information. The recent Ontario experience is instructive in this regard. An attempt at consensus development under the auspices of the Ontario Council of Health did not achieve expectations, in part because of the conflicting views of vested interests represented in the exercise and in part because of the difficulty of establishing and interpreting rigorous evaluation guidelines. Although recommendations were made for the case of diagnostic imaging (Ontario Council of Health, 1982), the broader exercise was terminated before the full series of conferences was completed. This, in large part, is the type of situation that motivates such efforts as this book on health technology assessment.

Thus the assembly and synthesis of information on health care technology in Canada still relies upon the efforts of individual researchers, health care providers, institutions, or government branches, working with what is essentially the literature from research journals and conference proceedings on health services. Although some in-house government synthesis may be done across the country, this is seldom issued publicly to prospective purchasers and utilizers of new technology.

Another element of the current approach to assembly and synthesis of information is continuing education programs for health professionals on new and emerging technologies. While diagnostic sub-specialty organizations have sponsored such programs for their members for many years, it is only recently that medical decision-making courses with an emphasis on the rational use of technological resources have been developed for front-line clinicians who initiate most decisions about the utilization of health care technology.

Policy Recommendation on the Production, Synthesis, and Dissemination of Information on Technology Assessment

A clear implication of the guidelines on technology assessment presented in Chapters 4 through 7, and of the messages in Chapter 8 on optimal facilities planning, is the need for timely and rigorous evaluation of major new health care technologies. To accomplish this, however, current methods of organizing and funding the production, synthesis, and dissemination of information on technology assessment must be improved.

Such information has a public good characteristic in that one person's use of the information does not reduce the amount available for use by others. Because the cost of producing new information may be high while the cost of disseminating it to users is relatively low, and because it is generally not possible to exclude those who do not pay for the information flow from using it, the tasks of producing and disseminating information are often best left to the public sector.

Similarly, each provincial ministry of health has a limited incentive to invest in evaluation of new technologies; while the province bears all of the cost, part of the benefits are captured by other provinces (and countries) who wait to use the results of studies done elsewhere. Thus the bulk of the responsibility for generating the information base on health technology assessment logically falls on the federal government (with international collaboration where possible).

As described above, however, there is fragmentation even within the efforts of the federal government. Although the recently intensified interest of the Medical Research Council in clinical trials is a positive step, it is unlikely to be sufficient. An important policy option is the creation of a National Health Technology Assessment Council (NHTAC).[2] It would be the responsibility of this agency to identify new and emerging technologies, select important technologies for rigorous assessment, fund scientifically sound clinical and economic evaluations, provide a liaison with international health technology evaluation efforts, disseminate results, and act as a general clearing house. Because the provincial ministries of health would be among the major consumers of the output of such an agency, their membership on the board of governors and/or participation in the funding of NHTAC might be appropriate.

In the absence of such a national agency, as a second best alternative, serious consideration should be given to a federal-provincial advisory committee on health technology assessment similar to the one that has operated for health manpower. Another (not mutually exclusive) set of strategies would involve realignment of existing health research funding agencies to provide more support for health care technology assessment studies. Although these secondary suggestions are steps in the right direction, they would do little to overcome the fragmentation of research efforts or to provide for the synthesis and integration of knowledge.

While NHTAC would in part replace or supplement evaluation research funded by existing provincial and federal agencies, it would also be responsible for conducting more systematic and comprehensive investigation of new technologies than is undertaken currently. The creation of NHTAC as a mission-oriented research council with some in-house research capability, coupled with a primary reliance on extramural project and program research, should increase the productivity of assessment efforts. Because sound assessment would be the primary goal of the council, it should also be possible to hold it accountable.

It is always difficult to decide on the appropriate amount of funding for research. Technology assessment, which often involves the evaluation of results of earlier research, is no exception. It does appear likely, however, that we have been underfunding technology assessment, and the creation of NHTAC would be an important remedial step. Observers of the biomedical research scene in the United States have pointed out that "high technology" industries typically devote 5 to 10 per cent of their budgets to research and development.[3] Yet Chalmers (1981) reported that total private and public expenditures on medical research in the United States represent only 2 per cent, of total health care expenditures. In Canada the figure is closer to only 1 per cent, as shown in Table 11.1.

With respect to the evaluative research component of medical or health research, the largest publicly funded health research agency in the United States, the National Institutes of Health, spends only approximately 5 per cent of its budget on clinical trials, the primary evaluative strategy (Banta and Behney, 1981, p. 453). Fineberg and Hiatt (1979) have suggested that 1 per cent of total health care expenditures should be devoted to health technology assessment – an amount far greater than current expenditures and roughly equal to that for all health research in Canada rather than just evaluative research on health technology.

Among the responsibilities of NHTAC would be the identification of new and emerging technologies in order that important ones might be selected for evaluation and that evaluation results would be produced promptly, rather than after the diffusion process is well under way. Respected health researchers within the council would have to maintain contacts in major provincial and federal medical research agencies and other health care research establishments both inside and outside Canada to gather necessary information. Gathering information on new technologies from equipment manufacturers who have a proprietary interest in their innovations would be an even more formidable problem.

Once new technologies have been identified, selected, and evaluated, mechanisms are needed to promote the synthesis and dissemination of the evaluation results. These are very important steps. In the synthesis process, the ideal information set as specified in the guidelines will generally not be available in its entirety, and the gaps will need to be identified. In many cases less rigorous evidence and seasoned judgement may have to be applied to reach conclusions, while areas for additional investigation are flagged. As the reader may recall from Chapter 10, in the case of neonatal intensive care (NIC) the lack of numerous randomized controlled trials (RCTs) of the unit as a whole was "overcome" by an examination of both the RCT results on specific manoeuvers employed within such units and regional neonatal mortality trends before and after the introduction of NIC.

One potentially efficacious way to promote both synthesis and dissemination of evaluation results is the consensus development conference.[4] The type of conference envisaged here is based on prior circulation of commissioned synthesis reports by leading experts, followed by conference discussion and debate resulting in "state of the art" conclusions on the appropriate adoption and use of the new technology. Conference participants would include clinicians who use or would use the technology in question as well as methodologists in clinical trials, biostatistics, and economic evaluation.

The entire conference would be based on the presumption that the kinds of evidence called for by the guidelines are necessary, even if not sufficient, for a proper technology assessment. Another criterion for such conferences would be that, as frequently as possible, they be national rather than provincial, thereby promoting the nationwide dissemination of results after the conference and perhaps moderating the influence of vested interests. Furthermore, while the consensus development exercise should allow input from all affected parties, participants must bring their expertise to bear as individuals rather than as representatives of pressure groups, competing institutions, or corporations. Such a consensus development exercise might also be an appropriate forum for a consideration of the ethical and social implications of the adoption and use of the new technology.

Through the process of synthesis of the available evidence in the commissioned papers and conference discussions, and because of the participation at the conference of a "blue ribbon" panel of respected clinicians and methodologists, the results and recommendations of the conference would probably be influential both in promoting the use of effective new technologies and in delaying the diffusion of ineffective ones. The credibility that is likely to be produced by such a process will be a valuable asset in generating voluntary compliance and in justifying regulatory interventions (to be discussed below).

Other forums for consensus development include evaluation programs operated by professional associations, hospital associations, and ad hoc task forces.[5] In these cases it is again important that guidelines such as those proposed in this book be applied to assess the quality of the evidence.

An additional strategy for dissemination is continuing medical education for health professionals. By conveying the criteria for clinical and economic evaluation, and the results of specific applications of those criteria, the foundation for better decision making can be laid. The faculty of Health Sciences at McMaster University, for example, has provided evaluation workshops for health care professionals and administrators on issues including critical evaluation of evidence on diagnosis, therapy, clinical measurement, and economic evaluation.[6] This format of small group, problem-based learning is one approach to the dissemination of both methodological guidelines and the results of studies on specific technologies.

Most effective methods of behaviour modification involve performance assessment in one form or another (such as practice audit) and feedback. It is likely, however, that in the absence of incentives to use the information the effect will be relatively minor (Haynes et al., 1984). Unless the main actors in the system are both equipped and stimulated to utilize the information, improvements in adoption and utilization decisions are unlikely to be forthcoming.

While it is important to consider the inclusion of the guidelines criteria in continuing medical education programs, it may be equally if not more important to consider their inclusion in undergraduate medical school curricula. Because the pace of change of knowledge in medical practice is so rapid, it is important that health care professionals learn to be skeptical and discerning consumers of new information, that is, "learn how to learn." Isolated changes in the curricula in the absence of incentives to apply that knowledge, however, may not be very effective.

The strategies discussed above are designed to assemble and synthesize the information needed to produce better informed adoption and utilization decisions. Even if these strategies were vigorously implemented, however, an important informational deficiency would remain. Although the strategies would provide better information for decision making and would enhance the capability of key actors in the system to utilize that information, basic "performance" information would still be lacking. Therefore an additional task for NHTAC would be the establishment of a registry on the use of health care technologies (see Martin, 1980; Relman, 1980, 1982; and Bunker, Fowles, and Schaffarzick, 1982). This registry could provide an information base about the adoption and utilization of new technologies that would be very useful to both researchers and policy makers. As was noted in Chapter 3, the existing literature is particularly thin in the area of the determinants of the level of utilization of technologies. Systematic periodic surveys of the facilities and techniques in use and the frequency of use would contribute to remedying this deficiency.

The Current Policy Approach to Technology Assessment in Canada: The Allocation of Health Care Resources

Canadian policy on the diffusion and utilization of new technologies has relied upon rather blunt fiscal instruments that are made possible by the sole-source-funded nature of the Canadian health care system. As Evans (1983) has pointed out, attempts to ensure appropriate diffusion and utilization in Canada have been "top down and indirect" rather than "bottom up and explicit." A policy of global reimbursement control has been used, which has left individual institutions and providers to make decisions on health care technology adoption and utilization as part of their reaction to overall resource constraints. Although a situation in which new technologies are competing for constrained funds provides an excellent opportunity to apply evaluative information on dimensions such as those suggested in this book, it seems fair to say that these opportunities have been far from fully utilized.

Historically, the important global constraints have been provincial government limitation of operating budgets for hospitals, restrictions on the supply and location of new technology through separate approval of capital requests by hospitals, and to a limited extent the use of specific reimbursement schedules through decisions on what will and will not be reimbursable procedures for hospitals and other providers. These global constraints on diffusion and utilization have been made possible by the sole-source-funded nature of the Canadian health care system, which has allowed provincial governments to channel new technology mostly through the hospital sector where it is subject to budgetary control. It should be noted, however, that hospitals are increasingly turning to private fund-raising efforts of their own in an attempt to elude provincial budgetary controls. Whether this trend will become a significant threat to sole-source funding is an open question.

Two complementary policy thrusts that have supported governments' indirect approach to technology assessment have been manpower policy and the buffering of governmental structures. Manpower policy has assisted in the control of technology because the Canadian effort to restrict the growth of specialists relative to general practitioners has exerted a moderating influence on the servicing intensity of the health care system and on pressure for expansion of technology. In several provinces, intermediate levels of government, such as regional hospital districts or district health councils, have been used as a buffering mechanism to force negotiation and the assignment of priorities to regional requests to be handled by regional rather than by provincial-administrators. Of course, "end runs" are always possible, and, at least anecdotally, these apparently are on the increase as hospital administrators and influential providers go directly to ministers of health.

Broadly speaking, however, the Canadian approach to the overall control of health care expenditures has been very successful in the past, both absolutely and by international standards.[7] A more serious question is whether this success can be maintained, and whether the Canadian health care system can be adapted to meet changing conditions. There is increasing friction between government and providers/institutions about the funding and provision of health services in general and about the introduction and utilization of health care technology in particular. Although evidence mounts that we could be doing better in the efficient provision of health care, the expectation of providers and institutions appears to be that existing practice patterns should not be changed. Rather, more money should be found to finance them. The stage is thus set for ever more serious confrontations because government is increasingly unable or unwilling to increase the revenues it raises to finance the health care system, especially during periods of slow economic growth.

An important question, too infrequently addressed, is what is society willing to pay for expanded health services and the continuation of our unnecessarily high-cost style of health care? Questions such as this go beyond technology assessment to rational delivery of health care more generally.

Technology Assessment Placed in the Broader Context of Rationalizing the Health Care System

It is important to realize that the rational introduction and use of technology is interwoven with rational patterns of health care delivery more generally, that is, with patterns that attempt to obtain value for money within a set of complex and sometimes conflicting policy goals such as universal access, equity, the provision of service capacity on the basis of need, and expenditure control.

Any consideration of policies to deal with technology must therefore be part of an overall consideration of policies for rationalizing health care delivery. These policies must affect the decisions of health care providers about what services are needed, in what amounts, and how the services should be produced. Herein lies a problem in the current structure of health care delivery in Canada. As noted by Bird (1981), Evans (1982), and Stoddart (1985), there is a fundamental structural problem in the organization and financing of Canadian health care delivery. The current system is characterized by a critical absence of linkage between the authority to make decisions about the provision of specific services and the financial responsibility to pay for those services. This situation has not occurred by accident. It represents a conscious decision by governments in Canada to adopt a "pay the bills" philosophy and to avoid interfering in the practice of medicine itself or in decisions about the utilization of specific services.

Incentives for efficient health care delivery are noticeably absent; in fact, they are perverse. Fee-for-service reimbursement of physicians encourages a service-intensive practice style. Community physicians have no incentive to make efficient use of hospitals and their technological capacity, because hospitals represent a "free" complementary capital input. Current methods of reimbursing hospitals, with few exceptions, neither encourage nor reward efficient institutions. Finally, consumers are protected by first-dollar universal insurance and are thereby insulated at the point of service (though not at tax time) from the cost of health care. Furthermore, even if consumers could ascertain and were interested in ascertaining cost-effective sources and styles of health care delivery, there currently exists no mechanism for such consumer decisions to affect health care providers. Stoddart (1985) has discussed the problem in detail. The important point here is that rationalization of health care delivery may require new organizational structures to accomplish the linkage between decision-making authority and financial responsibility.

The redesign of policies affecting health care delivery will not be an easy task, and several caveats should be kept in mind. First, it should be remembered that the system has performed quite well and has accomplished important public policy objectives, such as the provision of access to needed care regardless of income, the protection of consumers from financial catastrophe due to illness, and the provision of high-quality medical and hospital services on the basis of need. These are important accomplishments, and policies designed to improve the economic performance of the health care system, to be acceptable to Canadians, must recognize and preserve them.

Second, it should also be remembered that the problems which policies for rationalization address are not "people" problems but structural and/or incentive problems. The Canadian health care system is generally operated and utilized by dedicated providers and conscientious consumers; however, there are no incentives for these individuals to align their own goals and interests with a broader public interest in a cost-effective delivery system.

Third, policies must acknowledge that providers have significant discretion over the utilization process because of the asymmetry of information between consumers and providers as to what services exist, which services are appropriate for which conditions, and what those services will accomplish. Policies for rationalization of health care delivery, including the utilization of technology, must affect directly or indirectly the decisions of individual providers.

Fourth, policies must recognize that medicine is less precise than we often think.[8] Even among conscientious, well-trained health care providers, a range of practice styles exists for accomplishing the same therapeutic results. There are, therefore, both high-cost and low-cost styles of care. The challenge for policy is to promote the utilization of low-cost but still effective styles of practice.

Fifth, policies will have to recognize that fundamental concepts such as "need" and "health" are not well defined. It will always be possible, therefore, to provide more "needed" services or to expand health care activity in the hope of improving health. The nature of the health care sector is such that it will not place boundaries on itself. Nevertheless, although policy decisions may be difficult and may not always be made rationally, policies about health care delivery imply placing bounds on the provision of health care.

For all of these reasons, the linkage between decision making authority and financial responsibility in health care delivery generally, and in the control of technology in particular, must be accomplished carefully and thoughtfully. The main message for those concerned with technology assessment is that, although there are some specific policy options that can be addressed to technology assessment in isolation, and others that might warrant consideration within the existing system of organization and financing, the most significant and far-reaching policy options will likely involve changes in organization and financing that go beyond technology policy itself to affect basic decisions in the process of providing and utilizing all health care. Technology is but one input into the utilization process. If there are mechanisms and policies that rationalize the process, they will simultaneously encourage the rational use of technology.

In the following two sections policy options are addressed first within the current structure of organization and financing and then under circumstances requiring significant changes to that structure. In each section it is assumed that sufficient knowledge and evidence on the dimensions of safety, efficacy, effectiveness, and efficiency exist. The policy options therefore have to do with the goals of appropriate introduction and appropriate utilization of technology.

A few general comments may be in order. First, the options should not be viewed as mutually exclusive options, because various combinations of them might be considered. Indeed, it is likely that policy development must occur on several fronts simultaneously if the current performance of the health care system is to be improved. Second, the prescription for change is always easier than its execution. The operational and behavioural impediments to the implementation of these policies have not yet been addressed, and space does not permit detailed discussion of these issues. Policy changes will not occur overnight, and it would be naive to believe that it will occur without friction and attempts to block or derail it. Nonetheless, the directions indicated in these sections are overdue and must soon be acted on to have an effect in the decades to come. Third, the policy options discussed here are to a large extent "generic." Different and more specific strategies may be appropriate in the case of a particular technology. The policy milieu may also vary with the specific technology being considered. Finally, policies may be implemented by government or non-government organizations, including educational and private sector institutions.

Policy Options within the Existing Structure of Health Care Delivery

More Stringent Global Reimbursement Constraints

The Canadian policy of global reimbursement control has been relatively effective in the past; however, one course of action open to policy makers is to enact even more stringent measures. One example of this would be to hold fast on decisions to make no allowance in hospital operating budgets for the operating costs of any new equipment purchased with funds raised privately or acquired through philanthropy.[9] A second example would be the steadfast refusal of provincial governments and funding agencies in the future to "pick up the tab" for hospital deficits as they have done in the past. Provincial governments might take a particularly tough stance on deficits with hospitals that are deemed not to have given due consideration to evidence or guidelines in the acquisition of technology. While the main advantage of this course of action would be that hospitals would be allowed to adopt only those technologies that have a proven record of effectiveness, the main disadvantage would clearly be the increased confrontation between governments and institutions – and the political risk accompanying it – that such a policy would produce.

A variant on these options would be to explore the development of a stable (in real terms) multi-period capital budget for each hospital, separate from its operating budget, out of which the hospital would be expected to finance its adoption of new programs and acquisition of new equipment. This capital budget would include both start-up and operating costs for new activities.[10] As old activities were dropped or as previously purchased equipment was not replaced, additional funds would become available for the initiation of new programs. New acquisitions and programs would all be competing for limited funds through the same annual capital budgeting process rather than being examined serially. Decisions would initially be made within individual hospitals level in a decentralized fashion rather than by the ministry. An important component of the feasibility of implementing such a proposal would be the prior agreement of health professionals and hospital administrators on the criteria to be used in choosing among proposed projects.

More Direct Intervention in Hospital Decision Making

This policy line might be pursued in several ways. First, provincial government grants to hospitals for technology acquisition might be made conditional on the use of the technology according to certain clinical algorithms or protocols, with accompanying requirements for monitoring and reporting. The purpose of such protocols, which would be based on existing research evidence or expert consensus, would be to restrict the acquisition and use of technologies to those situations for which the technologies satisfied guidelines of need, effectiveness, and efficiency.

Under a more aggressive policy, certain classes of decisions about technology might be removed entirely from the authority of hospital administrators and clinical decision makers. Although it might be viewed as impossible for government to prohibit the acquisition of technologies by specific institutions not owned by government, the control of hospitals in the Canadian health care system clearly does rest with government. The ownership/control issue in Canadian hospitals could be the subject of a monograph in itself. Suffice it to say that examples exist in British Columbia and in Ontario of the provincial government's funding leverage and legislative mandate allowing for the direct, temporary takeover of at least two major hospitals "in the public interest." Direct public management of Canadian hospitals would undoubtedly be an extreme and highly unpopular reaction to the need to manage technology. Moreover, given that decisions about technology are only one subset of the decisions required in a complex institutional environment, it is not clear that government management would in general be superior. Nevertheless, it is the end point on the spectrum of more direct intervention in hospital decision making.

Selective Use of Reimbursement Schedules

The nature of reimbursement schedules for medical and hospital services in Canada already exerts one check on the proliferation of technology. As Evans (1983) has pointed out, new techniques are not initially represented by an item in fee or benefit schedules payable to physicians, but rather are introduced through the hospital budget process. Even when new techniques become established, however, restrictions are normally placed on who may be reimbursed for specific services. Hospitals are normally reimbursed for the technical component of the procedure, while physicians (either hospital-based or community-based) are reimbursed for their professional, often interpretive, services. Thus, although private diagnostic and laboratory services exist in Canada, most technological services are channelled through the hospital sector wherein opportunities for control exist.

Yet even in the current situation there may be scope for adjustment in the relative rates of reimbursement for certain services in order to influence their utilization in aggregate. This would involve a reduction, relative to other services, in reimbursement rates for those services deemed by provincial funding agencies to be overutilized, based on guidelines such as those discussed in this book. Marginally effective technologies might receive relatively low rates of reimbursement while cost-effective technologies used for an approved list of indications might receive more generous reimbursements.

The advantage of this policy is that it would align the economic interest of providers and institutions with the social interest in more efficient use of technology. Furthermore, evidence exists to support the view that health care providers do respond to such economic incentives.[11] The main disadvantage, however, is that this policy instrument is difficult to deploy with precision. While low reimbursement rates may be appropriate for unproven or marginally effective services (although in these cases it is questionable whether the services should be reimbursable at all), they seem less appropriate for procedures used at frequencies, or for conditions, beyond those proven effective. Lowering the reimbursement rate for a procedure renders its performance less financially attractive not only in the "target" situations but in all situations, including those where the procedure may be both necessary and effective.

It is also important to note at this point that the selective use of reimbursement schedules, like the preceding two options, would create a much greater administrative burden for provincial health ministries. One advantage of the current policy approach is that it economizes on the detailed information about physician practice and hospital operation that is required by payment agencies. The design of more powerful management-information systems would therefore be a necessary adjunct to the policy directions mentioned thus far.

Education and Peer Review Strategies

A broad range of activities are possible here and may be undertaken at the initiative of both government and non-government organizations. The question remains, however, whether any educational intervention would be sufficient to change behaviour in the absence of incentives to do so, and indeed in the presence of powerful incentives not to do so.[12]

In the context of continuing education and peer review there is considerable scope for audits of utilization and peer pressure to be brought to bear upon inappropriate under- or over-servicing. In a recent review of research into the effectiveness of continuing education for physicians, Haynes and colleagues (1984) found that some forms of continuing education did improve practitioner performance; however, the effective strategies generally involved direct practice audit and in-depth clinical retraining, both of which are much more expensive than the usual form of continuing education (didactic events with limited time involvement and high student-to-teacher ratios). Further, all of the studies of continuing education that have been rigorously executed to date have focused primarily on treatment interventions rather than on diagnostic tests; thus their application to technology use is limited at best. Finally, few scientifically rigorous assessments of continuing education to date have included an evaluation of cost-effectiveness or its congeners. Therefore, while it is clear that practitioners could improve their performance in the utilization of health care technology, and while it is possible that continuing education may help to achieve this end, it remains to be shown that continuing education itself is cost effective. This is an important research issue in the search for validated means of improving the dissemination and optimum use of health care technology.

A related strategy is the use of day-to-day monitoring of selected diagnostic tests and therapies. Approaches analogous to those used to deter antibiotic abuse could be adopted. McGowan and Finland (1974) describe a program in which physicians who ordered particularly expensive and toxic antibiotics were required to contact an infectious disease consultant before the drug would be released. Whereas audits are performed retrospectively, this strategy is one of prospective review. Although their study was not as well controlled as one would like, the dramatic differences in antibiotic usage they demonstrated suggest that this strategy may be an effective one. Such programs have not been run, however, for more than a few years. The reasons for this are not clear but may include antagonism on the part of the medical staff toward what they perceive as an intrusion into their practice, or fatigue on the part of the consultant whose job it is to make the inquiries.

More Formalized Procedures for Introduction and Dissemination

The highly formalized and tightly controlled process for introduction and dissemination of new technology in Sweden was described in Chapter 8. This process involves the designation of a few tertiary teaching and research centres as trial centres with specific protocols for research and evaluation. It also includes the prohibition of further dissemination of new technology outside these trial centres until initial evaluation results are known. It then progressively limits the dissemination of technologies down through various levels of the pyramid formed by the few tertiary teaching and research centres to the many local community hospitals. This process does go on to some extent in Canada; however, it is neither formalized nor consistent. One policy option open to provinces would be the formulation of specific rules and procedures for the phased dissemination of technology through trial centres. Like most of the other policy options within the existing structure of the health care system, this would require a toughness of political will that might be unrealistic to expect of Canadian government. Furthermore, it requires, on a case by case basis, the selection of the most appropriate trial centre(s), the adjudication of research and evaluation protocols, and difficult decisions about the appropriate time lag for the prohibition of further dissemination.

Policy Options Entailing Significant Changes to the Existing Structure of Health Care Delivery

Changes to Hospital Reimbursement Methods

Historically, Canadian hospitals have been paid on a cost-reimbursement basis. Although refinements have been made over time to the specific mechanisms, this system has persisted through the era of global budgets and even into the current era of prospectively negotiated global budgets.[13] Recently, more sophisticated techniques for hospital cost analysis have allowed the construction of average and marginal cost profiles for specific types of cases, sometimes indexed on several dimensions such as diagnosis and severity of illness. One very attractive policy option would appear to be the prospective reimbursement of hospitals on the basis of the amount and types of cases in their case load.[14]

Payment on a case basis, rather than through cost reimbursement, would provide powerful incentives for decision makers within hospitals to rationalize the input mix, especially with respect to technological inputs, in the provision of an episode of care for specific diagnoses or conditions. Romeo, Wagner, and Lee (1984) found that prospective reimbursement in the state of New York tended to speed the adoption of cost-saving technologies by hospitals while reducing the extent of adoption for cost-increasing technologies. Prospective reimbursement on a case basis would be attractive also in that it would require little interference in the practice of medicine, and it would leave decisions about the production of services in the hands of those who are presumably most knowledgeable – clinicians and hospital administrators.

Although prospective reimbursement on a case basis is attractive, especially with respect to attempts to rationalize the acquisition and use of technology, it is not without problems, as United States experience with diagnosis-related groups over the past few years has shown.[15] When receiving payment on a case basis, hospitals face incentives to reclassify cases into more remunerative diagnostic categories and may face incentives to refer extremely expensive cases to other institutions. The design of an appropriate case-rate structure is a challenging task. Nevertheless, this policy option could have a significant effect on the diffusion and utilization of health care technology.

Changes to the Internal Organization of Hospitals

The preceding policy option was intended to accomplish the linkage between decision-making authority and financial responsibility at the level of the hospital as an institution. Analogously, there is the need to accomplish the same linkage inside the hospital. Although the role for government initiative in this area is severely limited, hospital administrators themselves might consider (perhaps in response to changes in reimbursement methods) a different internal organization. One type of internal organization is that of clinical budgeting as outlined by Coles, Davison, and Wickings (1976), Wickings (1983), and Wickings and colleagues (1983). Clinical budgeting is a system in which clinical teams and units are given specific budgets, within a system of rules and incentives and are made financially responsible internally for their decisions in the provision of care. Such a change would require significant adjustments to the internal organization of Canadian hospitals; it is included here as an example of a change that would align the interests of individual decision makers with the broader institutional and governmental interest in the rational use of technology.

Changes to Incentives for Physicians Outside Hospitals

Just as the clinical department budget is designed to link financial responsibility with clinical decision-making authority within the hospital, health maintenance organizations (in the United States) and health service organizations (in Ontario) play a similar role in the provision of primary care. Under these delivery models, persons with the clinical and administrative authority to make decisions about the provision of services also have a direct financial stake in the effectiveness and efficiency of such services.[16]

Existing evidence suggests that, compared with traditional fee-for-service practice, HMOs and HSOs deliver equivalent health care with much less hospital utilization and consequently lower overall cost. Physicians practising within these organizations, however, also appear to make somewhat heavier use of ambulatory diagnostic services.[17] Overall, policy movement to such models might improve the utilization of technology because physicians would face incentives to eliminate inefficacious or cost-ineffective utilization of all types.

In comparisons of the performance of HMOs and HSOs to that of fee-for-service primary care physicians, it should be recognized that the significant differences may include more than the incentive structure faced by the physician. There is probably a process of self-selection occurring. Physicians who enjoy or tolerate more intense peer review, and/or who prefer certain practice "styles," may be more likely to join HSO practices. There may also be differences in the patient populations being served. Thus, generalizations from HMO/HSO performance should be made carefully. Even with these caveats, however, HSO-style practices, perhaps within a pluralistic setting including fee-for-service practices and community health centres (Stoddart and Seldon, 1984), appear to provide another attractive strategy for the promotion of the rational utilization of health care technologies in Canada.

Policies to Affect the Nature of Innovation

Preceding sections have dealt with the means to promote the appropriate diffusion and utilization of health care technology; some observers have suggested, however, that technology policy should include efforts to affect the kinds of innovations supplied. Just as granting agencies could be directed to fund more evaluative studies, they could also be instructed to weight the potential costs and benefits of basic versus applied research in making their allocations of research funds. If research on new forms of diagnostic imaging, for example, is deemed to have few benefits at the margin, research support for these programs would be curtailed. As the evidence cited in Chapter 2 on the relationship between the mortality and morbidity burden of a disease and its research funding makes clear, these considerations are already taken into account to some extent. Although the situation might be improved by expanding the concept of burden of disease to take quality of life more fully into account it is unlikely that such efforts to affect the nature of the stream of innovations will significantly alter the effect on the cost of new health care technologies. This is especially true in a "small" country like Canada where many of the innovations will have originated elsewhere.

There are other important reservations about a policy of attempting to direct the supply of innovations. First, a great deal of research is conducted in the private sector and would thereby be insulated from the policy thrust. Second, the kinds of innovations that are produced depend not only on the direction mandated but also on the supply of knowledge – the foundation provided by basic research. Enormous demand exists for a simple immunization against a disease like cancer, and the potential social pay-off of such an innovation is well known. We have no such immunization, however, because the disease has not, at least as yet, yielded to this approach.

A policy of direction has more relevance for applied research, but is perhaps better realized through actions to affect the demand for innovations rather than their supply.[18] If technologies are routinely subjected to rigorous clinical and economic evaluation and if the results of those evaluations are used in adoption and utilization decisions, the nature of the demand by providers of health care for innovations will be profoundly altered. In that case suppliers of innovations in both the private and public sectors will respond to the new market conditions by self-screening and selection of their innovations and by altering the thrust of their development efforts.[19]

Conclusions

A large menu of options exists, many of which are unlikely to work in isolation; instead they complement one other. Bold initiatives to generate and disseminate information on technology assessment require changes in the regulatory environment or incentive structure for their full potential to be realized. Within the Canadian health care system it is likely that there will be gradual "reforms" along some of the avenues described. Because decision making about the provision of services in the existing health care system is highly decentralized and diffuse, those policy changes that align financial and clinical responsibilities contain the greatest potential to contribute to the promotion of the rational use of health care technologies. Thus, although they are more difficult to implement, policies to encourage alternative health care delivery organizations, such as HSOs, or to alter hospital budgeting processes – internally, regionally, and provincially – can help to achieve appropriate patterns of technology diffusion and utilization by fostering the efficient delivery of all health services.

One implication of almost all of the policy options discussed is the need for increased research, planning, and administrative capability throughout the health care system. Most provincial ministries of health, for example, would lack the expertise, even if they had the mandate, for optimal facilities planning of the type described in Chapter 8. Few funding agencies possess the sophistication of equipment or analysts to implement quickly a case-type hospital reimbursement system. Few hospitals have internal accounting systems geared toward rationalizing the input mix in the delivery of an episode of care. Finally, there does not appear to exist anywhere in the Canadian system a group of experts with sufficient knowledge about technology assessment to inform policy making, to manage international linkages, and guide to consensus forming exercises. The time for technology assessment policy initiatives has come.

Notes

1. Although a specific technological advance, such as nuclear magnetic resonance, may spur governments to make a special effort to assemble available information, such cases are the exception rather than the rule and are normally performed on an ad hoc basis. Furthermore, federal and provincial governments may designate one or more persons to monitor developments in the health technology area, but ongoing and comprehensive exercises in data assembly and synthesis require considerably more resources and expertise than are normally devoted to these tasks.

2. Although the proposed NHTAC is not patterned after any specific existing institution, its closest antecedent is the moribund National Center for Health Care Technologies (NCHCT) in the United States. The proposed model for NHTAC combines functions of the U.S. NCHCT and the Office of Medical Applications of the U.S. National Institutes of Health. A great deal of work on health technology assessment and policy has been done at the Office of Technology Assessment (OTA) of the U.S. Congress, an advisory group organized as a Congressional Staff Office in 1972. The activities of OTA were reflected in the 1978 legislation that created NCHCT within the Department of Health and Human Services. The 1981 budget, however, included no funds for NCHCT. For more detail on these agencies, see Banta and Behney (1981), Banta, Behney, and Willems (1981), and Blumenthal (1983); for an international perspective see Banta and Russell (1981). In 1984 the National Center for Health Services Research and Health Care Technology Assessment was created within the U.S. Department of Health and Human Services to make recommendations about which technologies should be covered under federally funded programs. Its mandate includes assessments of safety, effectiveness, and cost-effectiveness, and the legislation includes provision for the establishment of an advisory

body, a National Advisory Council on Health Care Technology Assessment; see Marwick (1984).

3. See *Report of the President's Biomedical Research Panel*, Appendix A (United States, 1976); p. 21. According to that source, research and development expenditures in the U.S. Department of Defense, for example, constitute approximately 10 per cent of the total budget. Medical research in Canada is discussed in Manga (1985).

4. For a description and discussion of the National Institutes of Health Consensus Development Conferences, see Perry (1979) and Perry and Kalberer (1980).

5. An example of evaluation and consensus development by a professional association of medical specialists is the Clinical Efficacy Assessment Project of the American College of Physicians. In 1983 this group planned to undertake evaluations of endoscopy, colonoscopy, cardiac pacemakers, biofeedback, and strategies for the diagnosis of breast cancer. A similar program is being undertaken by the American College of Cardiology. The American Medical Association's Diagnostic and Therapeutic Technology Assessment project focuses on the assessment of safety and efficacy. The Society of Obstetricians and Gynecologists of Canada and the Canadian College of Medical Geneticists both recently recommended that chorionic villi sampling, a new technique in prenatal diagnosis, be made available only within randomized clinical trials so that its safety and accuracy might be assessed before it becomes widely available. In the United States several private hospital chains and major insurance companies provide members (hospitals and insurance company subsidiaries, respectively) with evaluations and information on new devices and equipment. Finally, an example of an ad hoc task force is the Canadian Task Force on the Periodic Health Examination (1979).

6. See Department of Clinical Epidemiology and Biostatistics, McMaster University (1980, 1981, 1983, 1984), Haynes et al. (1984), and Chambers, Stoddart, and Sullivan (1983). See also Boudreau et al. (1973) and Baskin et al. (1980).

7. See, for example, Canada (1983) and Barer and Evans (1985) for comparative data on health expenditures in

Canada and the United States. For a selection of international experiences with health care expenditure control, see McLachlan and Maynard (1982).

8. For an interesting and important discussion of "what medical care is and isn't" see Enthoven (1980), Chap. 1.

9. A variant of this policy would require hospitals that propose capital acquisitions financed by private donations to satisfy guidelines on need, effectiveness, and efficiency.

10. This idea came from discussions with Professor Theodore R. Marmor, chairman, Center for Health Studies, Yale University. Although the idea is suggested here in the context of a hospital's capital budget, the proposed capital budgeting plan could instead be implemented for an entire region, at least in jurisdictions where regional or district bodies possess fiscal authority.

11. Evans (1983) notes several instances in Canada in which the rate of performance of particular procedures was apparently altered by changes in the financial attractiveness of the procedure to physicians. More generally, physician sensitivity to economic incentives in the practice of medicine has been a prominent theme in the health economics research literature. For an early example, see Monsma (1970); for a more recent one, see Rice (1983).

12. This theme is discussed by Schroeder et al. (1984) in their report on a major, two-year, prospective, controlled study of the effect of physician education on both the utilization of specific procedures and on total hospital costs at a large university hospital in the United States. Although the educational interventions (weekly lectures and audit with feedback) did show a positive effect on some test-ordering patterns of physicians, the authors concluded that "in the absence of other cost containing incentives, physician education alone is not an effective hospital cost containment strategy", (p. 225).

13. Some provinces have experimented with policies that allow hospitals to retain a portion of any underspending (although the future budget base is typically reduced), thereby creating some incentive for efficiency, dependent upon the proportion of the underspending that is retainable and the

restrictions on its use. The most notable current example of such a reimbursement policy is the Business Oriented New Development (BOND) program in Ontario. In this system, underspending is fully retained by hospitals, with few restrictions on the redeployment of saved funds. Thus hospitals that successfully rationalize facility provision and/or utilization patterns stand to benefit and may generate reserves from which to finance new programs. The BOND system also encourages hospitals to create additional revenue through certain approved commercial activities, such as charges for private rooms, office rentals to physicians, and the operation of parking garages and shopping facilities. This aspect of the system, while potentially an important source of revenue for new programs and equipment, does little, however, to encourage rationalization of existing facilities, equipment, and utilization patterns.

14. For review and explanation of these techniques, see Horn and Schumacher (1979) and Barer (1982). Horn (1983) has recently discussed the application of these techniques to prospective reimbursement policies in a Canadian context.

15. See, for example, Davies and Westfall (1983), Neuhauser (1983), Anderson and Steinberg (1984), and Gonnella, Hornbrook, and Lewis (1984).

16. For a description of the HMO program in the United States, and a review of evidence on the efficiency of HMOs, see Luft (1981). The characteristics of HSOs in Ontario are outlined in Ontario (1982). Although there are numerous variants of the basic models, the central characteristic is the commitment on the part of the practice to provide a defined set of medical services to a specified population during a given period in return for a fixed sum of money.

17. In addition to Luft (1981), see Hastings et al. (1973) and Manning et al. (1984).

18. The importance of the basic versus applied research distinction is illustrated by Moloney and Rogers (1979) who report that of the 529 important publications on cardiovascular and pulmonary disease over the previous thirty years, more than 40 per cent came from investigators

who had no direct interest in the clinical problems and whose goal was simply to expand general knowledge.

19. For evidence on the responsiveness of technological change to the demand for innovations, as well as the importance of considering the supply of technical change (that is, the effects of the stock of knowledge on the costs of innovating), see Rosenberg (1972).

References

Anderson, Gerard, and Steinberg, Earl. 1984. "To Buy or Not to Buy: Technology Acquisition under Prospective Payment." *New England Journal of Medicine* 311(3): 182-185.

Banta, H. David, and Behney, Clyde J. 1981. "Policy Formulation and Technology Assessment." *Milbank Memorial Fund Quarterly* 59(3): 445-479.

Banta, H. David; Behney, Clyde J.; Willems, Jane Sisk. 1981. *Toward Rational Technology in Medicine: Considerations for Health Policy*. New York: Springer.

Banta, H. David, and Russell, Louise B. 1981. "Policies toward Medical Technology: An International Review." *International Journal of Health Services* 11(4): 631-652.

Barer, Morris L. 1982. "Case Mix Adjustment in Hospital Cost Analysis." *Journal of Health Economics* 1(1): 53-80.

Barer, Morris L., and Evans, Robert G. 1985. "Riding North on a South-Bound Horse? Expenditures, Prices, Utilization and Incomes in the Canadian Health Care System." In *Medicare at Maturity: Achievements, Lessons, and Challenges*, ed. Robert G. Evans and Greg L. Stoddart. Toronto: Irwin, forthcoming.

Baskin, Marjorie; Levesque, Louise; MacPherson, Alexander S.; Poole, E.; Sackett, David L. 1980. "Canada's Health Care Evaluation Seminars: An Epilogue and Evaluation." *Canadian Journal of Public Health* 71(5): 321-327.

Bird, Richard M. 1981. *Commentaries on the Hall Report*, 1-25. Toronto: Ontario Economic Council.

Blumenthal, David. 1983. "Federal Policy toward Health Care Technology: The Case of the National Center." *Milbank Memorial Fund Quarterly* 61(4): 584-613.

Boudreau, T.J.; Last, J.M.; Poole, P.E.; Sackett, D.L. 1973. "The National Health Grant Health Care Evaluation Seminars." *Canadian Journal of Public Health* 64(3): 260-264.

Bunker, John P.; Fowles, Jinnet, Schaffarzick, Ralph. 1982. "Evaluation of Medical-Technology and Strategies." *New England Journal of Medicine* 306: 620-624, 687-692.

Canada. Health and Welfare. 1981. *National Health Expenditures in Canada, 1970-1979.* Ottawa: Health and Welfare Canada.

Canada. Health and Welfare. 1983. *National Health Expenditures in Canada, 1970-1982.* Ottawa: Health and Welfare Canada.

Canadian Task Force on the Periodic Health Examination. 1979. "The Periodic Health Examination." *Canadian Medical Association Journal* 121(9): 1193-1254.

Chalmers, Thomas C. 1981. "The Clinical Trial." *Milbank Memorial Fund Quarterly* 59(3): 324-339.

Chambers, Larry W.; Stoddart, Greg L.; Sullivan, Brian. 1983. "Continuing Education for Health Professionals and Administrators: Workshops on Becoming a Critical User of Health Care Research." *Canadian Journal of Public Health* 74(1): 29-34.

Coles, James; Davison, Alan; Wickings, Iden. 1976. "Allocating Budgets to Wards: An Experiment." *The Hospital and Health Services Review* 72(9): 309-312.

Davies, R.H., and Westfall, G. 1983. "Reimbursement Under DRG's: Implementation in New Jersey." *Health Services Research* 18(2): 233-247.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1980. "Clinical Disagreement: I. How Often it Occurs and Why" and "II. How to Avoid it and

How to Learn from One's Mistakes." *Canadian Medical Association Journal* 123(6): 499-505 and (7): 613-617.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1981. "How to Read Clinical Journals: I. Why to Read Them and How to Start Reading Them Critically," "II. To Learn about a Diagnostic Test," "III. To Learn the Clinical Course and Prognosis of Disease," "IV. To Determine Etiology or Causation," and "V. To Distinguish Useful from Useless or Even Harmful Therapy." *Canadian Medical Association Journal* 124(5): 555-558, (6): 703-710, (7): 869-872, (8): 985-990, (9): 1156-1162.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1983. "Clinical Epidemiology Rounds: Interpretation of Diagnostic Data: 1. How to Do It with Pictures," "2. How to Do It with a Simple Table (Part A)," "3. Hot to Do It with a Simple Table (Part B)," "4. How to Do It with a More Complex Table," "5. How to Do It with Simple Maths," and "6. How to Do It with More Complex Maths." *Canadian Medical Association Journal* 129(5): 429-432, (6): 559-564, (7): 705-710, (8): 832-835, (9): 947-954, (10): 1093-1098.

Department of Clinical Epidemiology and Biostatistics, McMaster University. 1984. "Clinical Epidemiology Rounds: How to Read Clinical Journals: VII. To Understand an Economic Evaluation (Part A and B)." *Canadian Medical Association Journal* 130(11): 1428-1434 and (12): 1542-1549.

Enthoven, Alain C. 1978. "Cutting Cost without Cutting Quality of Care." *New England Journal of Medicine* 298(22): 1224-1238.

Enthoven, Alain C. 1980. *Health Plan: The Only Practical Solution to the Soaring Costs of Medical Care.* Reading, Mass.: Addison-Wesley.

Evans, Robert G. 1982. "Health Care in Canada: Patterns of Funding and Regulation." In *The Public/Private Mix for Health,* ed. Gordon McLachlan and Alan Maynard, 369-424. London: Nuffield Provincial Hospitals Trust.

Evans, Robert G. 1983. "The Fiscal Management of Medical Technology: The Case of Canada." In *Resources for Health,* ed. H. David Banta, 178-199. New York: Praeger.

Fineberg, Harvey H., and Hiatt, Howard H. 1979. "Evaluation of Medical Practices: The Case for Technology Assessment." *New England Journal of Medicine* 301(20: 1086-1091.

Gonnella, J.S.; Hornbook, M.C.; Louis D.Z. 1984. "Staging of Disease: A Case-Mix Measurement." *Journal of the American Medical Association* 251(5): 637-644.

Hastings, John F.; Mott, F.D.; Barclay, A.B.; Hewitt, D. 1973. "Prepaid Group Practice in Sault Ste. Marie, Ontario: Part I. Analysis of Utilization Records." *Medical Care* 11(2): 91-103.

Haynes, R. Brian; Davis, David A.; McKibbon, Ann; Tugwell, Peter. 1984. "A Critical Appraisal of the Efficacy of Continuing Medical Education." *Journal of the American Medical Association* 251(1): 61-64.

Horn, Susan D. 1983. "Controlling Clinical Expenditures." In *Proceedings of the Conference on Managing Canada's Health Care System,* 67-85. Saskatoon, August 30-31, 1983. Toronto: The Financial Post.

Horn, Susan D., and Schumacher, D.N. 1979. "An Analysis of Case-Mix Complexity Using Information Theory and Diagnosis Related Grouping." *Medical Care* 17(4): 382-389.

Luft, Harold S. 1981. *Health Maintenance Organizations: Dimensions of Performance.* New York: John Wiley.

Manga, Pran. 1985. "Economic and Ethical Determinants of the Allocation of Resources for Medical Research." University of Ottawa Faculty of Administration Working Paper 85-29.

Manning, W.G.; Leibowitz, A.; Goldberg, G.A.; Rogers, W.H.; Newhouse, J.P. 1984. "A Controlled Trial of the Effect of a Prepaid Group Practice on Use of Services." *New England Journal of Medicine* 310(23): 1505-1510.

Martin, David L. 1980. "Health Technology: Tiger or Pussycat?" Ottawa: Health Services Directorate, Department of National Health and Welfare. Mimeo.

Marwick, Charles. 1984. "Legislation Expands Federal Role in Medical Technology Assessment." *Journal of the American Medical Association* 252(23): 3235-3237.

McGowan, John F., and Finland, Maxwell. 1974. "Usage of Antibiotics in a General Hospital: Effects of Requiring Justification." *Journal of Infectious Diseases* 130(2): 165-168.

McLachlan, Gordon, and Maynard, Alan, eds. 1982. *The Public/Private Mix for Health*. London: Nuffield Provincial Hospitals Trust.

Moloney, Thomas W., and Rogers, David E. 1979. "Medical Technology: A Different View of the Contentious Debate over Costs." *New England Journal of Medicine* 301(26): 1413-1419.

Monsma, George N. 1970. "Marginal Revenue and the Demand for Physicians' Services." In *Empirical Studies in Health Economics*, ed. Herbert E. Klarman, 145-160. Baltimore: Johns Hopkins Press.

Neuhauser, Duncan. 1983. "DRG's in Jersey." *Effective Health Care* 1(3): 153-154.

Ontario. Ministry of Health. 1982. *Report of the Task Force to Review Primary Health Care*. (Mustard Report). Toronto: Ministry of Health.

Ontario Council of Health. 1982. *Task Force on Diagnostic Imaging. Final Report*. Part II. Toronto: Ontario Council of Health.

Perry, Seymour. 1979. "NIH Panels Analyze Issues on Technology." *Hospitals* 53(11): 115-116.

Perry, Seymour, and Kalberer, John T., Jr. 1980. "The NIH Consensus-Development Program and the Assessment of Health-Care Technologies: The First Two Years." *New England Journal of Medicine* 303(3): 169-172.

Relman, Arnold S. 1980. "Assessment of Medical Practices: A Simple Proposal." *New England Journal of Medicine* 303: 153-154.

Relman, Arnold S. 1982. "An Institute for Health-Care Evaluation." *New England Journal of Medicine* 306: 669-670.

Rice, Thomas H. 1983. "The Impact of Changing Medicare Reimbursement Rates on Physician-Induced Demand." *Medical Care* 21(8): 803-815.

Romeo, Anthony A.; Wagner, Judith L.; Lee, Robert H. 1984. "Prospective Reimbursement and the Diffusion of New Technologies in Hospitals." *Journal of Health Economics* 3(1): 1-24.

Rosenberg, Nathan. 1972. *Technology and American Economic Growth*. New York: Harper & Row.

Schroeder, Steven A.; Myers, Lois P.; et al. 1984. "The Failure of Physician Education as a Cost Containment Strategy." *Journal of the American Medical Association* 252(2): 225-230.

Stoddart, Greg L. 1985. "Rationalizing the Health Care System." In *Ottawa and the Provinces: The Distribution of Money and Power*, ed. Thomas J. Courchene, David W. Conklin, and Gail Cook, vol. 2, 3-39. Toronto: Ontario Economic Council.

Stoddart, Greg L., and Seldon, James R. 1984. "Publicly Financed Competition in Canadian Health Care Delivery: A Proposed Alternative to Increased Regulation?" In *Proceedings of the Second Canadian Conference on Health Economics*, ed. J.A. Boan, 121-160. Regina: University of Regina.

United States. Department of Health, Education and Welfare. Public Health Service. 1976. *Report of the President's Biomedical Research Panel. Appendix A: The Place of Biomedical Science in Medicine and the State of the Science.* Washington, D.C.: U.S. Government Printing Office.

Wickings, Iden. 1983. "Controlling Clinical Expenditures." In *Proceedings of the Conference on Managing Canada's Health Care System*, 67-85, Saskatoon, August 30-31, 1983. Toronto: The Financial Post.

Wickings, Iden; Coles, James M.; Flux, Ray; Howard, Linda. 1983. "Review of Clinical Budgeting and Costing Experiments." *British Medical Journal*, 1: 575-578.

The Members of the Institute

Board of Directors
The Honourable John B. Aird, O.C., Q.C.
(Honorary Chairman)
Aird & Berlis, Toronto
The Honourable Robert L. Stanfield,
P.C., Q.C., (Chairman), Ottawa
Roger Charbonneau (Vice-Chairman)
Président du conseil d'administration
Banque Nationale de Paris (Canada)
Montréal
Dr. Robert Bandeen
President and Chief Executive Officer
Cluny Corporation, Toronto
L. I. Bell
Chief Executive Officer
Vancouver City Savings Credit Union
Louis A. Desrochers, Q.C.
McCuaig, Desrochers, Edmonton
Peter C. Dobell
Director, Parliamentary Centre for
Foreign Affairs and Foreign Trade
Ottawa
Dr. Rod Dobell
President, The Institute for Research
on Public Policy, Victoria
Dr. Regis Duffy
President
Diagnostic Chemicals Ltd.
Charlottetown
Dr. James D. Fleck
Faculty of Management Studies
University of Toronto
Peter C. Godsoe
Vice Chairman of the Board
The Bank of Nova Scotia, Toronto
Dianne I. Hall
Senior Vice-President,
NOVA, AN ALBERTA
CORPORATION, Calgary
David Hennigar
Atlantic Regional Director
Burns Fry Limited, Halifax
Roland J. Lutes, C.A.
Clarkson Gordon, Montreal
Dr. Patrick O'Flaherty
Head, Department of English
Language and Literature, Memorial
University of Newfoundland, St. John's
Dr. Tom Pepper
Pepper Consultants Ltd., Saskatoon
Guy Roberge, C.R.
Conseil
Clarkson, Tétrault, Ottawa
Thérèse P. Sévigny
Vice-présidente à la communication
Société Radio-Canada, Ottawa
Robert K. Siddall
Senior Vice-President and Secretary
The Great-West Life Assurance
Company, Winnipeg
Marc-Adélard Tremblay, O.C.
Départément d'anthropologie
Université Laval, Sainte-Foy

President
Rod Dobell

Vice-President and
Secretary-Treasurer
Peter C. Dobell
Director, Parliamentary Centre
for Foreign Affairs and Foreign
Trade, Ottawa

Executive Committee
The Honourable Robert L. Stanfield
(Chairman)
Dr. Roger Charbonneau
(Vice-Chairman)
Peter C. Dobell
Dr. Rod Dobell
Dr. James D. Fleck
Roland J. Lutes

Investment Committee
Peter C. Godsoe, Chairman
Marcel Cazavan
Peter C. Dobell
Michael Koerner

Council of Trustees
Members at Large
Dr. Stefan Dupré (Chairman)
Department of Political Science
University of Toronto
Doris H. Anderson, O.C.
Toronto
Marie-Andrée Bertrand
École de criminologie
Université de Montréal
Dr. Roger A. Blais, O.C.
Directeur, R-D coopératifs
École Polytechnique de Montréal
George Cooper, Q.C.
McInnes, Cooper and Robertson
Halifax
James S. Cowan, Q.C.
Partner, Stewart, MacKeen & Covert, Halifax
V. Edward Daughney
President, First City Trust Company, Vancouver
Dr. Wendy Dobson
President
C.D. Howe Institute, Toronto
Dr. H.E. Duckworth, O.C.
Chancellor
University of Winnipeg
Marc Eliesen
Chairperson
Manitoba Energy Authority
Winnipeg
Emery Fanjoy
Secretary, Council of Maritime Premiers, Halifax
Dr. Allan K. Gillmore
Executive Director, Association of Universities and Colleges of Canada, Ottawa
Dr. Donald Glendenning
President, Holland College
Charlottetown
Margaret C. Harris
Past President, The National Council of Women of Canada
Saskatoon
Michael Hicks
Institute of Public Administration of Canada, Ottawa
Richard W. Johnston
President, Spencer Stuart & Associates, Toronto
Dr. Leon Katz, O.C.
Saskatoon
Dr. David Leighton
Director, National Centre for Management Research and Development, University of Western Ontario, London
Terrence Mactaggart
Managing Director
Sound Linked Data Inc.
Mississauga
Judith Maxwell
Chairman, Economic Council of Canada, Vanier
Claude Morin
École nationale d'administration publique, Québec
Milan Nastich
Canadian General Investments Limited, Toronto
Professor William A. W. Neilson
Dean, Faculty of Law
University of Victoria
Roderick C. Nolan, P.Eng.
President
Neill & Gunter Limited, Fredericton
Robert J. Olivero
United Nations Secretariat
New York
Gordon F. Osbaldeston
Senior Fellow, School of Business Administration, University of Western Ontario, London
Garnet T. Page, O.C.
Calgary
Dr. Gilles Paquet
Dean, Faculty of Administration
University of Ottawa
Dr. K. George Pedersen
President, University of Western Ontario, London
Professor Marilyn L. Pilkington
Osgoode Hall Law School, Toronto
Dr. Stuart L. Smith
Chairman Science Council of Canada, Ottawa
Eldon D. Thompson
President, Telesat, Vanier

Dr. Israel Unger
Deean of Science
University of New Brunswick
Fredericton
Philip Vineberg, O.C., Q.C.
Phillips & Vineberg, Montreal
Dr. Norman Wagner
President and Vice-Chancellor
University of Calgary
Ida Wasacase, C.M.
Winnipeg
Mr. Ronald L. Watts
Department of Political Studies
Queen's University, Kingston
Dr. R. Sherman Weaver
Director, Alberta Environmental
Centre, Vegreville
Dr. Blossom Wigdor
Director, Program in Gerontology
University of Toronto

Government Representatives
Herb Clarke, Newfoundland
Joseph H. Clarke, Nova Scotia
George Ford, Manitoba
George de Rappard, Alberta
Hershell Ezrin, Ontario
Honourable Lowell Murray, Canada
John H. Parker, Northwest Territories
Roger Burke, Prince Edward Island
Norman Riddell, Saskatchewan
Christian Dufour, Québec
Norman Spector, British Columbia
Eloise Spitzer, Yukon
Barry Toole, New Brunswick

Institute Management

Rod Dobell	President
Peter Dobell	Vice-President and Secretary-Treasurer

Yvon Gasse	Director, Small & Medium-Sized Business Program
Barbara L. Hodgins	Director, Western Resources Program
John Langford	Director, Governability Research Program
Barry Lesser	Director, Information Society Studies Program
Shirley Seward	Director, Studies in Social Policy
Frank Stone	Director, International Economics Program

Parker Staples	Director, Financial Services
Donald Wilson	Director, Communications
Louis Vagianos	Director, Publishing Services

Tom Kent	Editor, *Policy Options Politiques*

Fellows- and Scholars-in-Residence:

Edgar Gallant	Fellow-in-Residence
Tom Kent	Fellow-in-Residence
Eric Kierans	Fellow-in-Residence
Jean-Luc Pepin	Fellow-in-Residence
Gordon Robertson	Fellow-in-Residence
Richard Brown	Scholar-in-Residence

Publications Available - September 1986

Order Address:

The Institute for Research on Public Policy
P.O. Box 3670 South
Halifax, Nova Scotia
B3J 3K6

Leroy O. Stone & Claude Marceau	*Canadian Population Trends and Public Policy Through the 1980s.* 1977 $4.00
Raymond Breton	*The Canadian Condition: A Guide to Research in Public Policy.* 1977 $2.95
Raymond Breton	*Une orientation de la recherche politique dans le contexte canadien.* 1977 $2.95
J.W. Rowley & W.T. Stanbury (eds.)	*Competition Policy in Canada: Stage II, Bill C-13.* 1978 $12.95
C.F. Smart & W.T. Stanbury (eds.)	*Studies on Crisis Management.* 1978 $9.95
W.T. Stanbury (ed.)	*Studies on Regulation in Canada.* 1978 $9.95
Michael Hudson	*Canada in the New Monetary Order: Borrow? Devalue? Restructure!* 1978 $6.95
David K. Foot (ed.)	*Public Employment and Compensation in Canada: Myths and Realities.* 1978 $10.95
Raymond Breton & Gail Grant Akian	*Urban Institutions and People of Indian Ancestry: Suggestions for Research.* 1979 $3.00
Thomas H. Atkinson	*Trends in Life Satisfaction Among Canadians, 1968-1977.* 1979 $3.00
W.E. Cundiff & Mado Reid (eds.)	*Issues in Canadian/U.S. Transborder Computer Data Flows.* 1979 $6.50
Meyer W. Bucovetsky (ed.)	*Studies in Public Employment and Compensation in Canada.* 1979 $14.95
Richard French & André Béliveau	*The RCMP and the Management of National Security.* 1979 $6.95
Richard French & André Béliveau	*La GRC et la gestion de la sécurité nationale.* 1979 $6.95

G. Bruce Doern & Allan M. Maslove (eds.)	*The Public Evaluation of Government Spending.* 1979 $10.95
Leroy O. Stone & Michael J. MacLean	*Future Income Prospects for Canada's Senior Citizens.* 1979 $7.95
Richard M. Bird	*The Growth of Public Employment in Canada.* 1979 $12.95
Richard J. Schultz	*Federalism and the Regulatory Process.* 1979 $1.50
Richard J. Schultz	*Le fédéralisme et le processus de réglementation.* 1979 $1.50
Lionel D. Feldman & Katherine A. Graham	*Bargaining for Cities, Municipalities and Intergovernmental Relations: An Assessment.* 1979 $10.95
Elliot J. Feldman & Neil Nevitte (eds.)	*The Future of North America: Canada, the United States, and Quebec Nationalism.* 1979 $7.95
David R. Protheroe	*Imports and Politics: Trade Decision Making in Canada, 1968-1979.* 1980 $8.95
G. Bruce Doern	*Government Intervention in the Canadian Nuclear Industry.* 1980 $8.95
G. Bruce Doern & Robert W. Morrison (eds.)	*Canadian Nuclear Policies.* 1980 $14.95
Allan M. Maslove & Gene Swimmer	*Wage Controls in Canada: 1975-78: A Study of Public Decision Making.* 1980 $11.95
T. Gregory Kane	*Consumers and the Regulators: Intervention in the Federal Regulatory Process.* 1980 $10.95
Réjean Lachapelle & Jacques Henripin	*La situation démolinguistique au Canada: évolution passée et prospective.* 1980 $24.95
Albert Breton & Anthony Scott	*The Design of Federations.* 1980 $6.95
A.R. Bailey & D.G. Hull	*The Way Out: A More Revenue-Dependent Public Sector and How It Might Revitalize the Process of Governing.* 1980 $6.95
David R. Harvey	*Christmas Turkey or Prairie Vulture? An Economic Analysis of the Crow's Nest Pass Grain Rates.* 1980 $10.95
Donald G. Cartwright	*Official Language Populations in Canada: Patterns and Contacts.* 1980 $4.95
Richard M. Bird	*Taxing Corporations.* 1980 $6.95
Leroy O. Stone & Susan Fletcher	*A Profile of Canada's Older Population.* 1980 $7.95
Peter N. Nemetz (ed.)	*Resource Policy: International Perspectives.* 1980 $18.95

Keith A.J. Hay (ed.)	*Canadian Perspectives on Economic Relations With Japan.* 1980 $18.95
Dhiru Patel	*Dealing With Interracial Conflict: Policy Alternatives.* 1980 $5.95
Raymond Breton & Gail Grant	*La langue de travail au Québec : synthèse de la recherche sur la rencontre de deux langues.* 1981 $10.95
Diane Vanasse	*L'évolution de la population scolaire du Québec.* 1981 $12.95
David M. Cameron (ed.)	*Regionalism and Supranationalism: Challenges and Alternatives to the Nation-State in Canada and Europe.* 1981 $9.95
Heather Menzies	*Women and the Chip: Case Studies of the Effects of Information on Employment in Canada.* 1981 $8.95
H.V. Kroeker (ed.)	*Sovereign People or Sovereign Governments.* 1981 $12.95
Peter Aucoin (ed.)	*The Politics and Management of Restraint in Government.* 1981 $17.95
Nicole S. Morgan	*Nowhere to Go? Possible Consequences of the Demographic Imbalance in Decision-Making Groups of the Federal Public Service.* 1981 $8.95
Nicole S. Morgan	*Où aller? Les conséquences prévisibles des déséquilibres démographiques chez les groupes de décision de la fonction publique fédérale.* 1981 $8.95
Raymond Breton, Jeffrey G. Reitz & Victor F. Valentine	*Les frontières culturelles et la cohésion du Canada.* 1981 $18.95
Peter N. Nemetz (ed.)	*Energy Crisis: Policy Response.* 1981 $10.95
James Gillies	*Where Business Fails.* 1981 $9.95
Allan Tupper & G. Bruce Doern (eds.)	*Public Corporations and Public Policy in Canada.* 1981 $16.95
Réjean Lachapelle & Jacques Henripin	*The Demolinguistic Situation in Canada: Past Trends and Future Prospects.* 1982 $24.95
Irving Brecher	*Canada's Competition Policy Revisited: Some New Thoughts on an Old Story.* 1982 $3.00
Ian McAllister	*Regional Development and the European Community: A Canadian Perspective.* 1982 $13.95
Donald J. Daly	*Canada in an Uncertain World Economic Environment.* 1982 $3.00
W.T. Stanbury & Fred Thompson	*Regulatory Reform in Canada.* 1982 $7.95

Robert J. Buchan, C. Christopher Johnston, T. Gregory Kane, Barry Lesser, Richard J. Schultz & W.T. Stanbury	*Telecommunications Regulation and the Constitution.* 1982 $18.95
Rodney de C. Grey	*United States Trade Policy Legislation: A Canadian View.* 1982 $7.95
John Quinn & Philip Slayton (eds.)	*Non-Tariff Barriers After the Tokyo Round.* 1982 $17.95
Stanley M. Beck & Ivan Bernier (eds.)	*Canada and the New Constitution: The Unfinished Agenda.* 2 vols. 1983 $10.95 (set)
R. Brian Woodrow & Kenneth B. Woodside (eds.)	*The Introduction of Pay-TV in Canada: Issues and Implications.* 1983 $14.95
E.P. Weeks & L. Mazany	*The Future of the Atlantic Fisheries.* 1983 $5.00
Douglas D. Purvis (ed.), assisted by Frances Chambers	*The Canadian Balance of Payments: Perspectives and Policy Issues.* 1983 $24.95
Roy A. Matthews	*Canada and the "Little Dragons": An Analysis of Economic Developments in Hong Kong, Taiwan, and South Korea and the Challenge/Opportunity They Present for Canadian Interests in the 1980s.* 1983 $11.95
Charles Pearson & Gerry Salembier	*Trade, Employment, and Adjustment.* 1983 $5.00
Steven Globerman	*Cultural Regulation in Canada.* 1983 $11.95
F.R. Flatters & R.G. Lipsey	*Common Ground for the Canadian Common Market.* 1983 $5.00
Frank Bunn, assisted by U. Domb, D. Huntley, H. Mills, H. Silverstein	*Oceans from Space: Towards the Management of Our Coastal Zones.* 1983 $5.00
C.D. Shearing & P.C. Stenning	*Private Security and Private Justice: The Challenge of the 80s.* 1983 $5.00
Jacob Finkelman & Shirley B. Goldenberg	*Collective Bargaining in the Public Service: The Federal Experience in Canada.* 2 vols. 1983 $29.95 (set)
Gail Grant	*The Concrete Reserve: Corporate Programs for Indians in the Urban Work Place.* 1983 $5.00
Owen Adams & Russell Wilkins	*Healthfulness of Life.* 1983 $8.00
Yoshi Tsurumi with Rebecca R. Tsurumi	*Sogoshosha: Engines of Export-Based Growth.* (Revised Edition). 1984 $10.95
Raymond Breton & Gail Grant (eds.)	*The Dynamics of Government Programs for Urban Indians in the Prairie Provinces.* 1984 $19.95

Frank Stone	*Canada, The GATT and the International Trade System.* 1984 $15.00
Pierre Sauvé	*Private Bank Lending and Developing-Country Debt.* 1984 $10.00
Mark Thompson & Gene Swimmer	*Conflict or Compromise: The Future of Public Sector Industrial Relations.* 1984 $15.00
Samuel Wex	*Instead of FIRA: Autonomy for Canadian Subsidiaries?* 1984 $8.00
R.J. Wonnacott	*Selected New Developments in International Trade Theory.* 1984 $7.00
R.J. Wonnacott	*Aggressive US Reciprocity Evaluated with a New Analytical Approach to Trade Conflicts.* 1984 $8.00
Richard W. Wright	*Japanese Business in Canada: The Elusive Alliance.* 1984 $12.00
Paul K. Gorecki & W.T. Stanbury	*The Objectives of Canadian Competition Policy, 1888-1983.* 1984 $15.00
Michael Hart	*Some Thoughts on Canada-United States Sectoral Free Trade.* 1985 $7.00
J. Peter Meekison Roy J. Romanow & William D. Moull	*Origins and Meaning of Section 92A: The 1982 Constitutional Amendment on Resources.* 1985 $10.00
Conference Papers	*Canada and International Trade. Volume One: Major Issues of Canadian Trade Policy. Volume Two: Canada and the Pacific Rim.* 1985 $25.00 (set)
A.E. Safarian	*Foreign Direct Investment: A Survey of Canadian Research.* 1985 $8.00
Joseph R. D'Cruz & James D. Fleck	*Canada Can Compete! Strategic Management of the Canadian Industrial Portfolio.* 1985 $18.00
Barry Lesser & Louis Vagianos	*Computer Communications and the Mass Market in Canada.* 1985 $10.00
W.R. Hines	*Trade Policy Making in Canada: Are We Doing it Right?* 1985 $10.00
Bertrand Nadeau	*Britain's Entry into the European Economic Community and its Effect on Canada's Agricultural Exports.* 1985 $10.00
Paul B. Huber	*Promoting Timber Cropping: Policies Toward Non-Industrial Forest Owners in New Brunswick.* 1985 $10.00
Gordon Robertson	*Northern Provinces: A Mistaken Goal. 1985* $8.00

Petr Hanel	*La technologie et les exportations canadiennes du matériel pour la filière bois-papier.* 1985 $20.00
Russel M. Wills, Steven Globerman & Peter J. Booth	*Software Policies for Growth and Export.* 1986 $15.00
Marc Malone	*Une place pour le Québec au Canada.* 1986 $20.00
A. R. Dobell & S. H. Mansbridge	*The Social Policy Process in Canada.* 1986 $8.00
William D. Shipman (ed.)	*Trade and Investment Across the Northeast Boundary: Quebec, the Atlantic Provinces, and New England.* 1986 $20.00
Nicole Morgan	*Implosion: An Analysis of the Growth of the Federal Public Service in Canada (1945-1985).* 1986 $20.00
Nicole Morgan	*Implosion: analyse de la croissance de la Fonction publique fédérale canadienne (1945-1985).* 1986 $20.00
William A.W. Neilson & Chad Gaffield (eds.)	*Universities in Crisis: A Mediaeval Institution in the Twenty-first Century.* 1986 $20.00
Fred Wien	*Rebuilding the Economic Base of Indian Communities: The Micmac in Nova Scotia.* 1986 $20.00
D.M. Daly & D.C. MacCharles	*Canadian Manufactured Exports: Constraints and Opportunities.* 1986 $20.00
Gerald d'Amboise, Yvon Gasse & Rob Dainow	*The Smaller, Independent Manufacturer: 12 Quebec Case Studies.* 1986 $20.00
David J. Roy & Maurice A.M. de Wachter	*The Life Technologies and Public Policy.* 1986. $20.00
David Feeny, Gordon Guyatt & Peter Tugwell (eds.)	*Health Care Technology: Effectiveness, Efficiency, and Public Policy.* 1986 $20.00

The Institute for Research on Public Policy
L'Institut de recherches politiques

A national, independent, research organization
Un organisme de recherche national et indépendant

Créé en 1972, l'Institut de recherches politiques est un organisme national dont l'indépendance et l'autonomie sont assurés grâce a des revenus provenant d'un fonds de dotation auquel souscrivent les gouvernements fédéral et provinciaux ainsi que le secteur privé. L'Institut obtient en outre des subventions et des contrats des gouvernements, des compagnies et des fondations afin de réaliser certains projets de recherche.

La raison d'être de l'Institut est triple :

- Servir de catalyseur au sein de la collectivité nationale en favorisant un débat public éclairé sur les principales questions d'intérêt général.
- Stimuler la participation de tous les éléments de la collectivité nationale à l'élaboration de la politique d'État.
- Trouver des solutions réalisables aux importants problèmes d'ordre politique afin de contribuer à l'élaboration d'une saine politique d'État.

Un Conseil d'administration, chargé de la décision, et une Commission de direction, responsable d'éclairer le Conseil sur l'orientation de la recherche, assurent la direction de l'Institut. L'administration courante des politiques, des programmes et du personnel relève du président.

L'Institut fonctionne de manière décentralisée et retient les services de chercheurs en divers points du Canada, s'assurant ainsi que toutes les régions contribuent aux recherches.

L'Institut cherche à favoriser, dans la mesure du possible, la compréhension et la discussion publiques des questions d'envergure nationale, controversées ou non. Il publie les conclusions de ses recherches avec clarté et impartialité. Les recommandations ou les conclusions énoncées dans les publications de l'Institut sont strictement celles de l'auteur et n'engagent aucunement le conseil d'administration, la commission de direction ou les bailleurs de fonds.

Le président assume la responsabilité ultime de publier un manuscrit au nom de l'Institut. Il jouit à cette fin des conseils du personnel de l'Institut et de critiques de l'extérieur quant à l'exactitude et l'objectivité du texte. Ne sont publiés que les textes qui traitent de façon compétente d'un sujet digne de la réflexion du public.

Les publications de l'Institut paraissent dans la langue de l'auteur et sont accompagnées d'un abrégé dans les langues officielles du Canada.

HEALTH CARE TECHNOLOGY: Effectiveness, Efficiency, and Public Policy

Edited by
David Feeny, Gordon Guyatt, and Peter Tugwell

New health care technologies are often incorporated into the health care system without proper evaluation of their effectiveness or their efficiency. Dissatisfaction may develop as a technology fails to meet expectations; it falls into disuse and eventually is replaced by yet another new and incompletely evaluated technology.

To improve the performance of the health care system, the authors of this book present a comprehensive set of guidelines for the clinical and economic evaluation of health care technologies. They establish the kinds of evidence needed to conclude that a new technology is safe, effective and efficient.

In addition, the authors propose the creation of a federal National Health Technology Assessment Council with provincial government participation and a mandate to produce and disseminate the evaluations of new technologies. They encourage the promotion of a more rational health care delivery system in which providers have the incentive to use the results of evaluations. To that end, they explore policy options for changes in the organization of the system.

The book is written for a broad audience: health care professionals, health policy makers, and the consumers of health care services—the public. Through their analyses and their proposals, the authors have contributed to productive public policy debate on the issues in assessing health care technology.

ISBN 0-88645-038-1 PRICE: $20.00

The Institute for Research on Public Policy
L'Institut de recherches politiques

A national, independent, research organization
Un organisme de recherche national et indépendant